PROJECT MUSE®

The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change

Patrice Speeter Beddor, Andries W. Coetzee, Will Styler, Kevin B. McGowan, Julie E. Boland

Language, Volume 94, Number 4, December 2018, pp. 931-968 (Article)

Published by Linguistic Society of America
DOI: *https://doi.org/10.1353/lan.2018.0051*

For additional information about this article
https://muse.jhu.edu/article/711824

Access provided at 5 Oct 2019 10:44 GMT from Bayerische

THE TIME COURSE OF INDIVIDUALS' PERCEPTION OF COARTICULATORY INFORMATION IS LINKED TO THEIR PRODUCTION: IMPLICATIONS FOR SOUND CHANGE

Patrice Speeter Beddor
University of Michigan

Andries W. Coetzee
University of Michigan and *North-West University*

Will Styler
University of California, San Diego

Kevin B. McGowan
University of Kentucky

Julie E. Boland
University of Michigan

Understanding the relation between speech production and perception is foundational to phonetic theory, and is similarly central to theories of the phonetics of sound change. For sound changes that are arguably perceptually motivated, it is particularly important to establish that an individual listener's selective attention—for example, to the redundant information afforded by coarticulation—is reflected in that individual's own productions. This study reports the results of a pair of experiments designed to test the hypothesis that individuals who produce more consistent and extensive coarticulation will attend to that information especially closely in perception. The production experiment used nasal airflow to measure the time course of participants' coarticulatory vowel nasalization; the perception experiment used an eye-tracking paradigm to measure the time course of those same participants' attention to coarticulated nasality. Results showed that a speaker's coarticulatory patterns predicted, to some degree, that individual's perception, thereby supporting the hypothesis: participants who produced earlier onset of coarticulatory nasalization were, as listeners, more efficient users of nasality as that information unfolded over time. Thus, an individual's perception of coarticulated speech is made public through their productions.*

Keywords: individual differences, coarticulation, sound change, nasalization, speech production, speech perception

1. Introduction. This study investigates how individual speakers produce, and how those same language users perceive, the dynamics of coarticulation. The relation between speech production and perception has been a foundational issue for phonetic theory for half a century. An overarching goal of phonetic theory is to explain the principles of human speech production and perception that underlie effective transmission of a linguistic message: speakers produce articulations with acoustic consequences that convey the planned linguistic information to listeners. This effective transmission requires sufficient equivalence, or parity, between the forms of speaking and the forms of listening (Liberman & Whalen 2000). Consequently, theories of speech production and perception seek to determine the nature—for example, the linguistically relevant acoustic-auditory properties or gestural events (see Diehl et al. 2004 and Fowler 2007 for reviews)—of these shared forms of produced and perceived speech. For some of these theoretical approaches, the stipulation of shared forms both follows from the broad requirement of sufficient equivalence and extends to the forms produced and per-

* We thank Kerby Shedden for advice on statistical modeling, Skye Huerta and Karen Tan for assistance with data coding, and Anthony Brasher for help with pilot data collection. This work has greatly benefited from the comments of members of several audiences, especially those at ICPhS 2015, the 2017 Fourth Workshop on Sound Change, and the University of Michigan Phonetics-Phonology Research Group. Given that the current editor of *Language* is a coauthor of this article, the editorial process was anonymously managed by a guest editor. We express our appreciation to the guest editor for their professional handling of the manuscript, and to three anonymous referees for their valuable comments. This material is based on work supported by the National Science Foundation under Grant No. BCS-1348150 to Patrice Beddor and Andries Coetzee; any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Printed with the permission of Patrice Speeter Beddor, Andries W. Coetzee, Will Styler, Kevin B. McGowan, & Julie E. Boland. © 2018.

ceived by the individual language user. Most approaches to exemplar theory, for example, assume that a speaker's productions are drawn from exemplar clouds seeded by that individual's perceptual input; these models typically assume a perception-production loop in which the phonetic details of the perceived input are reflected in production (e.g. Pierrehumbert 2001).

A second domain in which the production-perception relation is a foundational issue is the study of sound change. Contemporary study of the phonetic underpinnings of sound change investigates how subphonemic phonetic variants might serve as a source of new sound patterns that spread through a speech community. Theoretical approaches that postulate that the listener initiates the change assume, implicitly or explicitly, a tight connection between the speech forms produced and perceived by a language user. A tight connection is inherent in the approach: if a listener's percept is to contribute to sound change, it must be publicly manifested. Arguably the most direct manifestation would be if there were a systematic relation between a listener's percept and that individual's productions. For example, in an influential account of the role of listeners in changes in which an originally predictable coarticulatory property (i.e. a property due to the overlap of articulatory gestures) becomes contrastive and the source of coarticulation is lost (as in, for example, the historical development of distinctive vowel nasalization), Ohala (1981) proposed that a listener may fail to detect the coarticulatory trigger (e.g. a nasal coda). That listener might therefore interpret the coarticulatory effects as an inherent property of the coarticulated portion of the signal (in this example, a nasalized vowel). The listener's interpretation is subsequently made public through the listener-turned-speaker's own productions. Although subsequent studies have proposed alternative accounts of perceptually motivated changes, most retain the assumption that the listener's innovation is mirrored in their production patterns (e.g. Lindblom et al. 1995, Harrington et al. 2008, Beddor 2009, Yu 2013).

This study tests whether an individual speaker-listener's articulatory repertoire—or grammar—predicts, in part, their perceptual grammar. Motivated by our conviction that the study of time-varying processes such as coarticulation requires time-sensitive measures, we study the time course of speakers' production of velum lowering in anticipation of an upcoming nasal consonant and the time course of these same participants' perceptual use of those anticipatory cues. We hypothesize that listeners who closely attend to the coarticulatory information, and who find that information especially useful in making linguistic decisions, will, as speakers, consistently and more extensively produce that information in their own speech. Although the broad goal of our research program is to address long-standing claims of theories of speech perception and theories of sound change, our main focus in this article is to address assumptions regarding the potential contributions of listeners-turned-speakers to sound change.

1.1. Individual variation in producing and perceiving coarticulated speech. We take as our starting point that speakers differ in their patterns of articulatory coordination and timing, and that listeners differ in their use of those patterns and in the decisions that they arrive at based on the time-varying acoustic input. Studies of the production and perception of gestural overlap illustrate these individual differences.

Although coarticulation is a necessary consequence of the temporal and spatial coordination of articulatory gestures, coarticulation is not exclusively automatic; rather, it is also planned by the speaker (Whalen 1990). This planning emerges in language-specific (e.g. Boyce 1990, Manuel 1990) and speaker-specific strategies of interarticulator coordination. Examples of the latter strategies are found in patterns of anticipatory lip, tongue-body, and velum position. Swedish speakers, for example, have been shown to produce two types of patterns of anticipatory lip movement during consonants preceding a rounded vowel: for some speakers, the longer the consonant sequence, the earlier the onset of lip movement; for others, lip rounding begins at a roughly constant interval before vowel onset irrespective of the length of the consonant sequence (Lubker & Gay 1982). Subsequent studies of the timing of anticipatory lip rounding have also reported individual strategies for speakers of Canadian French (Noiray et al. 2011), American English (Noiray et al. 2011), and Cantonese (Yu 2016). Speaker-specific patterns of anticipatory tongue-body position emerge in Grosvald's (2009) acoustic measures of the coarticulatory influences of a stressed vowel on preceding consonant-schwa sequences, which suggest that some English speakers begin to anticipate the tongue-body position for the stressed vowel as early as three syllables prior to the stressed syllable, whereas other speakers' schwa productions show no or only very small anticipatory influences (see also Magen 1997). Also for English, the temporal and spatial extent of vowel nasalization has been shown to differ across speakers, as demonstrated via aerodynamic (Cohn 1990:152, 177), kinematic (Krakow 1989:51–57), and acoustic (Beddor 2009) measures.

Along similar lines, perceptual findings demonstrate listener-specific patterns of attention to, or adjustment for, information regarding coarticulatory overlap. In one widely used perceptual paradigm, listeners identify members of, for instance, a consonant continuum embedded in vocalic contexts that differ in their coarticulatory influences on the consonant. Listeners' context-dependent judgments indicate that they adjust or compensate for the acoustic effects of coarticulation. However, listeners differ from each other in the magnitude of these compensatory adjustments. For example, anticipatory lip rounding for a rounded vowel lowers the frequencies of a preceding fricative, such that /s/ is acoustically more /ʃ/-like before, say, /u/ than /a/. Yu (2010) found that although, overall, English-speaking listeners adjusted for the rounding effects and reported hearing more /s/ (as opposed to /ʃ/) in a rounded context, some listeners systematically compensated more than others for this contextual influence (see also Mann & Repp 1980). Using eye-tracking techniques, Beddor, McGowan, Boland, Coetzee, and Brasher (2013) assessed not compensation, but rather the perceptual usefulness of coarticulation, and found that participants differed from each other in their use of coarticulatory cues to anticipate the remainder of a word. Such interlistener differences in attention to the informational properties of the input are not sporadic, but have been found to be consistent over time (Idemaru et al. 2012) and across tasks (Yu & Lee 2014).

That there are speaker-specific coarticulatory routines and listener-specific strategies for processing coarticulatory information has, unsurprisingly, led to investigation of the factors that might underlie these differences. Previous work has linked individual differences in production and perception to cognitive processing style (Yu 2010, 2016), executive-function capacity (Kong & Edwards 2016), and, more speculatively, social awareness of speech variation (Garrett & Johnson 2013). While we recognize the importance of these contributing factors, our approach investigates not the source of individual differences but rather the possibility that a given speaker's coarticulatory routines may, to some extent, predict that individual's perception of coarticulation.

1.2. The relation between individuals' production and perception repertoires. Foundational to the current investigation is whether a language user's perceptual processing is mirrored in that individual's productions. In recent years, several sets of studies have investigated the relation between perception and production of coarticulation for groups of language users (e.g. for groups differing in age or native language) and for individuals. A number of these studies have explored the process of back-vowel fronting whereby, in many speech varieties, the back vowel /u/ is coarticulatorily fronted in alveolar contexts (e.g. English *dude*). In the speech of younger speakers of Southern British English, /u/-fronting is well underway and has been generalized beyond the original alveolar context. These younger language users both produce /u/ with fewer coarticulatory influences (i.e. they tend to front /u/ even in nonfronting contexts) and perceptually adjust less for consonantal effects on /u/ than do older speakers (Harrington et al. 2008). Fronting of lax /ʊ/ is a more recent change in this same variety of English, yet here again the (albeit more complex) pattern is for greater fronting in both production and perception for younger participants (Kleber et al. 2012). By comparison, Kataoka's (2011) study of speaker-specific coarticulatory patterns for /u/-fronting in American English failed to show a significant link between degree of coarticulatory variability in a speaker's /u/ productions and that language user's perceptual compensation for coarticulatorily variable /u/ realizations.

Investigations of the production and perception of the coordination of other articulatory gestures also provide mixed results. For example, speakers of different languages differ in the spatiotemporal extent of vowel-to-vowel coarticulation and anticipatory vowel nasalization, and native-speaking listeners' perception appears to parallel production: the greater the produced coarticulation by a group of native speakers, the more that group compensates perceptually for those coarticulatory effects (Beddor & Krakow 1999, Beddor et al. 2002). However, particularly for vowel-to-vowel coarticulation, attempts to establish a perception-production link at the level of individual speaker-listeners have not succeeded. Although Grosvald found in his 2009 study that American English speakers' anticipation of an upcoming stressed vowel varied from almost no coarticulation to coarticulatory effects spanning three syllables, the individuals who produced greater coarticulation were not perceptually more sensitive to those effects. Subsequent work that measured participants' neural activity (mismatched negativity) when hearing coarticulated schwas also failed to find a correlation between these individuals' perception and production of anticipatory tongue-body coarticulation (Grosvald & Corina 2012). More successful in establishing a perception-production link are the results of Zellou's 2017 study of anticipatory nasal coarticulation in American English, which showed that individuals who produced more anticipatory nasalization compensated more (i.e. attributed more of the coarticulatory effects to the nasal consonant context) in a discrimination task. These same participants' productions, though, did not predict their perception of coarticulatory nasalization in a rating task that assessed listeners' judgments of relative vowel nasality.

A further set of studies testing whether a listener's perceptual processing of gestural timing is mirrored in that individual's productions has explored the relation between stop-consonant voicing and the fundamental frequency (f0) of a following vowel. Stop phonation gives rise to f0 'perturbations' on the vowel, resulting in a higher f0 after a voiceless than after a voiced stop. The magnitude and temporal extent of this f0 perturbation—and, indeed, the timing of voicing itself in relation to stop release (i.e. voice onset time or VOT)—are variable across speakers and are variably used by different listeners. In American English, where VOT is the primary and f0 a secondary source of information for stop voicing (e.g. Whalen et al. 1993), Shultz, Francis, and Llanos (2012) found the expected interparticipant variation in the production and perception of these two properties, but failed to find, for individual speaker-listeners, a correlation between their produced and perceived weights of these properties. In Seoul Korean, a relatively recent sound change has resulted in f0 being the primary information (rendering VOT secondary) for younger speakers for the contrast between aspirated and lax stops (e.g. Kang 2014, Kwon 2015). Yet here too these individuals' perceptual weighting of f0 and

VOT do not correlate with their produced weights (Schertz et al. 2015). In Afrikaans, in what is arguably an ongoing sound change, vocalic f0 is a prominent cue for stop voicing in the productions of all speakers. Older Afrikaans speakers are more likely than younger ones to produce a VOT distinction between phonologically voiced and voiceless stops and, correspondingly, they perceptually weight voicing more heavily than do younger participants. For individual speaker-listeners, this correlation is present as well, although only weakly so (Coetzee et al. 2018).

The picture that appears to be emerging for production and perception of coarticulated speech is that the differences between groups of speaker-listeners—older compared to younger participants or native speakers of one language compared to speakers of another—are large (and/or systematic) enough for the spatiotemporal magnitude of production of the targeted variant to predict perceptual adjustments for, or perceptual weights of, that variant. However, establishing a link between an individual speaker-listener's articulatory and perceptual repertoires or grammars has proved more elusive (albeit not entirely so; e.g. Zellou 2017). This difficulty at the individual level has multiple possible sources. One likely source is that, despite compelling theoretical reasons for postulating a perception-production link, there are, as well, compelling reasons for not expecting a perfect relation. Human interlocutors are necessarily flexible perceivers: they regularly interact with speakers whose production patterns differ from their own. Successful communication depends on perceptual adaptation to those patterns, and the overwhelming evidence is that perception is malleable across different speakers, speaking rates, and more (see Samuel 2011 for a review). However, precisely because perception adapts, production need not be similarly malleable; that is, speakers who do not accommodate to their listeners are nonetheless intelligible under many circumstances (Pardo 2012). Moreover, specifically from the perspective of coarticulation, if coarticulatory information is perceptually useful—that is, if listeners actively use this time-varying aspect of the signal to track what speakers are saying—we might expect listeners to attend to this information in the input even if it is not highly characteristic of their own speech patterns. For both of these reasons, we might expect speaker-listeners to exhibit greater flexibility in perception than production.

A perception-production mismatch might be especially likely if the targeted phenomenon involves a sound change in progress. Specifically, if some members of a speech community realize a contrast primarily with one property and other members realize it with another property, perceivers presumably will attend to both properties. For producers of the innovative property, this situation would be one in which perception lags behind production: these individuals perceptually rely on the conservative property yet typically do not produce it (see Pinget 2015 for discussion). By contrast, in sound changes in which a coarticulatory effect becomes independent of its source (e.g. context-independent back-vowel fronting in some varieties of English), there is evidence that, at the early stages of the change, the 'waning' of coarticulation is more advanced in perception than in production (Harrington et al. 2012, Kleber et al. 2012, Harrington et al. 2019). Thus, the extent to which perception and production align for individual speaker-listeners—and the nature of that alignment—may well depend on how stable the patterns of variation are within the speech community.

In this investigation, we consider whether the lack of correspondence between perception and production of coarticulation for individual speaker-listeners found in most previous studies might be, in part, a consequence of those studies' use of static measures to assess inherently dynamic processes. The theoretical orientation that underlies the current study of the time course of coarticulation is that production and perception are flexible, dynamic, and active processes. That production is not a sequence of static states (although discrete patterns may be present in dynamic movements; see e.g. Iskarous 2005) is perhaps evident and uncontroversial. In production, vocal-tract actions do not align with traditional segments. As we have discussed, the execution of, say, lip rounding for a vowel or velum lowering for a nasal consonant typically begins before the target vowel or consonant. In speech perception, though, theoretical approaches differ in their accounts of whether and how listeners attend to the time-varying information afforded by coarticulation (e.g. Strange et al. 1983, Lindblom 1990, Fowler 2006). Yet the preponderance of current evidence shows that listeners are sensitive to these phonetic details and that the lawful nature of coarticulatory (and other) variation informs perception. Thus, as the coarticulatorily structured acoustic signal unfolds in real time, listeners use that structure to inform their linguistic decisions. Coarticulatorily appropriate information, for example, speeds listeners' responses (Martin & Bunnell 1982, Fowler 2005), and inappropriate coarticulatory cues slow responses (Whalen 1984). Investigation of the dynamics of perception using the visual-world paradigm shows that perception evolves over the course of listeners' moment-by-moment processing of the acoustic input (Dahan et al. 2001). Study of the time-varying properties of produced and perceived coarticulation may thus better capture individual differences and, consequently, may provide a particularly rigorous test of possible production-perception links.

The experiments reported here investigate the time course of a listener's use of coarticulatory information against that language user's production of those same coarticulatory actions. This work is framed within the overarching perspective that a language user's perception and production repertoires are complexly related in ways that are mediated by wide-ranging (linguistic, social, psychological, and other) factors. The specific hypothesis regarding that relation—that individuals' perceptual attention to a targeted coarticulatory property correlates with the spatiotemporal extent of their production of that property—is motivated by the theoretical orientation we have just described. It is informed not only by the studies of coarticulation already reviewed, but also by relevant (noncoarticulatory) studies of imitation and phonetic accommodation. These studies inherently investigate the perception-production relation in that they test whether a participant's productions are influenced by what they have just heard, such as the heard speech of a model voice or of a conversation partner. Their results show that perception and production ARE interconnected: listeners-turned-speakers adjust their speech, over the course of the task, to be more similar to the phonetic patterns of the input speech (see Pardo 2012 for a review). Moreover, individual participants imitate or accommodate to different degrees, with their patterns being mediated by phonological (Mitterer & Ernestus 2008, Nielsen 2011, Kwon 2015) and social (Pardo 2006, Babel 2012) factors.

Our specific approach to studying the perception-production relation is also driven by our understanding of a major source of individual listener differences in perceiving coarticulated speech. A fundamental task of the listener is to arrive at the linguistic message being conveyed by the speaker. Coarticulation provides what is, in principle, redundant information for the listener; under many if not most circumstances, both the coarticulatory effect and the trigger of that effect—for example, a vowel produced with velum lowering followed by a nasal stop—structure the acoustic signal. Because multiple cues for a given speech contrast are (typically) reliably present, different weightings of this information can nonetheless result in successful perception. Thus, different listeners, even listeners with similar linguistic experiences, may arrive at different perceptual weights. Beddor (2009, 2012), for instance, had English-speaking listeners identify stimuli drawn from a *bet–bent* continuum in which the temporal extent of vowel nasalization and the duration of the nasal consonant orthogonally varied. Stimuli with a nasalized vowel but an extremely short [n] were unambiguously *bent* for some listeners but unambiguously *bet* for (a very few) others. Although such perceptual differences would likely be resolved in conversational settings, clearly some listeners assign more perceptual weight to certain aspects of the input than others do. The question being tested here is whether an individual's perceptual weights are preserved in their productions and thereby transmitted to their interlocutors.

1.3. PRODUCING AND PERCEIVING ANTICIPATORY NASALIZATION IN AMERICAN ENGLISH. The targeted coarticulatory property is nasalization in American English. Essential to our approach is that the time course of producing and perceiving this property differs across individuals. As we have already indicated, speakers differ in the extent to which the velum-lowering gesture aligns with the oral closure for a nasal consonant coda: for some speakers, velum lowering appears to be aligned with vowel onset (or even earlier, such as during the approximants of *rent* or *want*), whereas for others the evidence is suggestive of closer alignment with the onset of the oral closure (e.g. Beddor 2009, Zellou 2017). Perceptually, listeners not only are sensitive to the presence or absence of anticipatory vowel nasalization (Krakow et al. 1988, Fowler & Brown 2000, Flagg et al. 2006), but are also VARIABLY sensitive to the information (Beddor 2009, Zellou 2017). Most relevant is the Beddor et al. 2013 study in which eye movements were monitored as participants heard, for example, *bend* with anticipatory vowel nasalization and saw visual images of target *bend* and competitor *bed*. The study found that, when a standard delay in programming an eye movement is factored in, visual fixations on *bend* (or, more generally, on the target CVNC image) tended to begin before listeners heard the nasal consonant. However, individual participants in that study differed from each other in the time course of their fixations on the target and competitor images. Whereas, on average, most participants initially fixated on the CVNC image on the basis of the coarticulatory information, some listeners more reliably used the coarticulatory information—as shown by shorter target-fixation latencies—than did others.

Thus, production and perception of coarticulatory nasalization have been shown to have the requisite systematic variation across individuals. Nasalization is also relevant to the particular focus of this study on the potential contributions of listeners-turned-speakers to sound change in that coarticulatory nasalization has been the trigger for the sound change VN > Ṽ in the history of many of the world's languages. Roughly one quarter of the world's languages have phonemic nasal vowels, and in most of these languages these vowels evolved from earlier sequences of a vowel followed by a nasal consonant (e.g. Hajek 1997, 2013).

This study addresses the following questions with regard to nasalization in American English: (i) What is the time course of the aerodynamic consequences of the overlapping lowered-velum and oral-constriction gestures? (ii) What is the time course of perception of coarticulatory nasalization? (iii) For the individual language user, what is the relation between (i) and (ii)—that is, what is the relation between a listener's coarticulated productions and that language user's dynamic use of that coarticulatory information in determining what a speaker has said? Production is measured using aerodynamic methods; perception is assessed using the visual-world paradigm. In fact, because the Beddor et al. 2013 study that used this paradigm showed reliably different fixation patterns for different participants, the perception study replicates that earlier study with a new, larger group of participants from whom we also gathered airflow data. Comparable production and perception data were elicited from participants who produced and perceived CVC and CVNC stimuli that differed in coda voicing (e.g. *bet*, *bent*, *bed*, *bend*). Coda voicing was manipulated in the 2013 perception study because, in English, anticipatory vowel nasalization tends to be temporally more extensive in $VNC_{voiceless}$ than in VNC_{voiced} sequences (Malécot 1960, Cohn 1990:152, 175). We retained that study's aim of investigating listeners' use of these time-varying cues in real-time processing and extended it to the same participants' production of those cues. We predict that participants who systematically produce CVNC words with earlier onset of nasal airflow will also assign heavier perceptual weight to coarticulatory nasalization, and will therefore fixate the target CVNC image as soon as coarticulatory information becomes available to them, and in particular faster than participants who produce CVNC words with later onset of nasal airflow. At this stage of our research, however, we do not have expectations concerning whether the proposed link might have production or perception at its source (i.e. whether production patterns engender perceptual attention or the reverse; see §5.2 for discussion).

The prediction of this work that has especially notable implications for theories of sound change is that innovative listeners will also be innovative speakers. In the development of distinctive vowel nasalization (VN > Ṽ), the historical change requires that the originally redundant property become the contrastive one. In our approach, the innovative listeners who rely heavily on the predictable property of coarticulatory nasalization in perception are expected to be innovative speakers who produce that property consistently and extensively, possibly at the expense of the distinctive property (N). In adopting this characterization, we are not suggesting that American English vowel nasalization is undergoing change. However, as we discuss in §5, perception-production relations within what are presumably stable patterns of variation arguably have implications for less stable settings and, consequently, for theories of perceptually motivated sound changes.

2. Production: the time course of anticipatory nasal airflow. Data collection for this combined production and perception study was conducted in three sessions, with two perception sessions preceding a single production session. Here, however, because the individual differences in production will be used to predict the perception data, we present the results of the production experiment first.

The production experiment analyzes American English speakers' nasal airflow patterns in CVNC words and, where appropriate, NVN words. Our choice of nasal airflow over acoustic measures was motivated by our goal of assessing production of nasalization rather than its acoustic consequences—the latter being influenced by interactions between the coupled oral and nasal cavities and not monotonically related to velum position (e.g. Stevens 1998). Although nasal airflow is also only an indirect indicator of velic behavior, it is a commonly used noninvasive proxy for assessing velopharyngeal aperture (e.g. Delvaux et al. 2008, Shosted et al. 2012). It also provides more fine-grained temporal information.[1]

We expect that, consistent with some previous studies of vowel nasalization (e.g. Cohn 1990, Beddor 2009), anticipatory nasal airflow will begin earlier in the vowel when the final consonant of CVNC is voiceless (e.g. *sent*, *bent*) than when it is voiced

[1] A potential drawback of airflow data is that degree of nasal airflow is influenced by the configuration of the oral cavity, with increased oral resistance (e.g. due to higher compared to lower vowels) possibly leading to increased nasal airflow, independent of changes in velum position (e.g. Krakow & Huffman 1993). However, because all speakers in this study produced the same vowels in the same contexts, airflow variation due to changing oral configuration should be roughly constant across participants.

(*send*, *bend*). More directly relevant to this investigation is the prediction that systematic differences between speakers will emerge in the time course of anticipatory nasal airflow, with some speakers producing earlier onset of nasal flow and/or heavier nasal airflow early in the vowel of CVNC words than other speakers.

2.1. METHODS.

PARTICIPANTS. Forty-two undergraduates at the University of Michigan successfully completed the production and perception experiments. Additional participants were recruited as well, but were eliminated due to difficulties calibrating the eye-tracker for specific individuals (eleven individuals), poor mask fit resulting in airflow leakage (twenty-four individuals), failure to reach criterion on eye-tracking trials (one individual; see §3.1), or failure to complete all sessions (three individuals). A further eight participants were recorded, but their data were discarded because the small mask used to accommodate their smaller faces resulted in flow values that were not comparable to those from the larger (adult-sized) mask. Participants were native English speakers with normal or corrected-to-normal vision and no known hearing loss or reading difficulties. They were paid for participating in the three (two perception, one production) testing sessions.

STIMULI. Two word lists were created, one consisting of words having the structures CVC and CVNC, and the other of NVN words plus fillers. The CV(N)C words consisted of five sets of minimal quadruplets whose members differed in the presence of N and the voicing of the final C (*bet-bed-bent-bend*, *let-lead-lent-lend*, *set-said-scent-send*, *wet-wed-went-wend*, and *watt-wad-want-wand*), along with eleven additional words (*dent*, *feed*, *feet*, *fiend*, *hint*, *hit*, *lid*, *lint*, *pant*, *pond*, *pot*). The five sets of quadruplets are the words used in the eye-tracking task. In the second set, the NVN words were eleven monosyllabic items (*man*, *manned*, *mean*, *means*, *meant*, *men*, *mend*, *mince*, *mint*, *mom*, *non*) and five disyllabic items with stress on the NVN syllable (*demeaned*, *mambo*, *Mindy*, *monster*, *Nancy*). Of primary interest in this study of the time course of coarticulatory nasalization are the CVNC words, which were investigated in both the production and perception experiments. NVN words were included to provide additional coarticulatory information; CVC words were included as fillers.

Productions of the CV(N)C words were elicited with visual images presented on a computer screen. These black-and-white drawings, most of which were used in the eye-tracking task, are described in §3.1. NVN productions were elicited orthographically, both because many of them were not easily imageable and because we wanted to keep manageable the number of images with which participants needed to be familiar.

PROCEDURE. Airflow data were gathered from each participant in the first phase of the production task. Acoustic data were also gathered, but in a second phase of the task, because the airflow mask rendered an acoustic recording unusable for reliable spectral analysis. However, this study reports only the methods and results for the airflow data. The production session lasted about one hour and typically occurred within one week of the second perception session.

Participants were seated in front of a computer monitor. Prior to data collection, participants were refamiliarized with the labels for the images of the CV(N)C words. (Initial familiarization of all CV(N)C words was done during perception testing.) Participants were shown, in a self-paced procedure, each of the randomly ordered images, and they produced the labels aloud. They were required to produce each label correctly twice before moving on to the main task. An incorrect answer resulted in the correct label being shown on the screen and the word being reentered into the randomization.

In the airflow-collection phase of the task, participants positioned the hand-held airflow mask against their faces, with instructions to create a tight but comfortable seal. Nasal airflow was captured via the Glottal Enterprises Oral-Nasal Airflow system using a split oral-nasal silicone mask with mesh port covers and two PT-2E airflow capture transducers. Prior to each block of airflow data collection, each transducer was calibrated by pushing 140 ml of air through a calibration box attached to the transducer; air escaped through a vented-mesh port identical to those in the mask. This produced a known-volume pressure signal, which was then used to calculate a conversion factor that mathematically transforms the electrical pressure response of the transducer into the volume of air (in ml) passing through the mask.

Stimuli were presented using SR Research Experiment Builder software. Stimulus presentation was blocked by CV(N)C and NVN words. For CV(N)C words, participants saw two images on the monitor corresponding to target and competitor words (e.g. *feed*, *lint*), separated by a cross. For NVN words, participants saw two orthographic words rather than images. After 750 ms, the cross became an arrow, with a superimposed L(eft) or R(ight), pointing toward the corresponding (left or right) image. Participants then produced the target word in a carrier sentence. A correct production in a trial with a *feed* image and L arrow would be '*Feed* is on the left'. The experimenter confirmed successful production with a key press. If the participant produced an incorrect word or did not respond within two seconds, the correct label was shown, and the missed trial was reentered into the randomization. The next trial began approximately 750 ms after the key press or 2000 ms after a missed trial. The purpose of having participants produce target words in response to changing arrows and, in the CV(N)C trials, of having them respond to images rather than orthography was to make the task more engaging than the typical reading of a word list, thereby possibly increasing the naturalness of productions.

Breaks were given after every sixty-two words (five blocks) for CV(N)C trials and after every fifty words (three blocks) for NVN. Participants were encouraged to lower the mask during airflow breaks. For all participants, airflow collection for CV(N)C words preceded that of NVN words. Excluding any repeated trials, there were 310 randomized CV(N)C trials (31 words × 10 repetitions) and 150 randomized NVN trials ((16 NVN words + 9 fillers) × 6 repetitions), for a total of 460 airflow trials per participant.

Data analysis. Airflow signals were analyzed for all CVC, CVNC, and NVN words, with the exception of eight words (*dent*, *feet*, *hint*, *hit*, *pant*, *man*, *manned*, *Nancy*) that were excluded prior to coding to provide balance across the data set for final-consonant voicing and vowel quality. (For example, words with /æ/ were excluded due to /Cænd/ gaps in the data set.)

Nasal airflow during the vowel portion of each signal was measured, as was the duration of the vowel. For CVNC words, nasal consonant duration was also measured. Vowel boundaries were delimited for all words, as were nasal boundaries for CVNC words, using TextGrid annotations in Praat (Boersma & Weenink 2013). As illustrated in Figure 1 for a token of *scent*, segmentation was based on the nasal and oral waveforms, and on spectrograms that were created from the residual acoustic data captured by the airflow transducers. Signals were low-pass filtered below 5000 Hz (to remove extraneous acoustic information) and high-pass filtered above 40 Hz (to remove the nonacoustic airflow signal). Boundaries for vowel onset and offset were placed at the first and last, respectively, visually identifiable pitch pulses of the vowel and were based primarily on the oral waveform. N onset (in CVNC) was identical to vowel offset. N offset was determined largely on the basis of cessation of the signal in the nasal waveform.

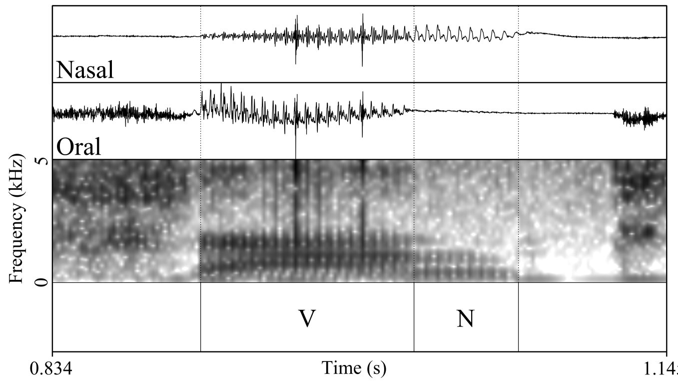

Figure 1. Nasal (top) and oral (middle) waveforms and speech spectrogram (bottom) for a token of *scent* (participant P55). See text for explanation of placement of V and N boundaries.

Despite precautions taken during recordings to minimize production errors, specific tokens were excluded from analysis due to speaker error (e.g. incorrect or disfluent target word, or carrier sentence not produced), experimenter error (e.g. early key press that cut off a portion of the target word), or an unanalyzable nasal waveform (due to mask slippage). On average, for each speaker, 92% of CVNC and 93% of NVN tokens were entered into the final calculations. (CVC productions were not entered into the statistical models, although we do represent these data in some of the figures below for the sake of comparison with CVNC and NVN.) For each speaker, at least 85% of that individual's CVNC and NVN productions were included, with three exceptions, for whom at least 73% were analyzable. However, even for the speaker with the fewest analyzable tokens, 175 CVNC and 72 NVN stimuli were included in the calculations.

2.2. Airflow results. In presenting the production results, we first briefly describe nasal airflow in all stimulus conditions (CVNC, NVN, CVC) to show that nasal airflow varied as expected (Fig. 2). Second, we present a detailed analysis of the CVNC conditions (voiced and voiceless; Fig. 3 and Table 1)—that is, of the conditions in which anticipatory nasal coarticulation in production will be compared with the perceptual use of such information in §4. In order to adequately describe the dynamic changes in volume of nasal airflow over time, the effects of time are modeled with three polynomial functions (B-splines, discussed below). Third, we conduct a functional principal component analysis to capture as much speaker variation as possible (Figs. 4 and 5). The second principal component from this analysis will be used to make speaker-level predictions about perception.

Figure 2 gives the average nasal airflow across normalized time for the CVC and CVNC conditions in voiced (left) and voiceless (right) contexts; NVN, which has no voicing context, is repeated in both panels. As expected, nasal airflow in the CVNC and NVN conditions increases over the latter portion of the vowel, consistent with anticipatory velum lowering, then rises steeply shortly before the onset of the nasal consonant (which corresponds to vowel offset). Also as expected, nasal airflow remains low and level across the vowel in the oral CVC condition.

To measure the time course of nasal airflow due to anticipatory nasalization in CVNC words, the airflow curves were sampled at twenty-five points, evenly spaced across the vowel's duration, for each CVNC token. Raw rather than normalized airflow values (in ml/sec) are collected and modeled here as the functional principal component analysis process described below effectively normalizes airflow-volume differences

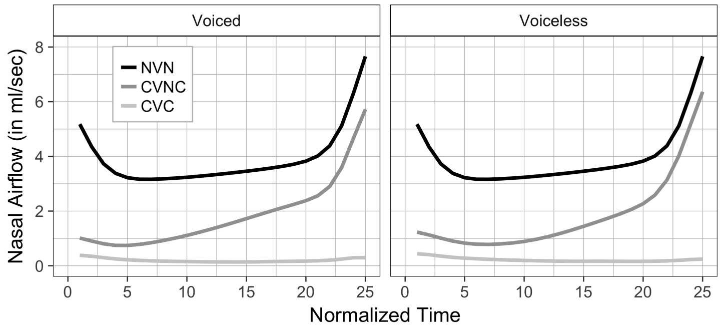

FIGURE 2. Mean nasal airflow across normalized time according to voicing (left, right panels) and nasality (line type) contexts.

across speakers. These points were entered into a linear mixed-effects model as the dependent variable using the lme4 package (Bates et al. 2014) in R (R Core Team 2013). *p*-values were obtained using the lmerTest package (Kuznetsova et al. 2015). As LMER models are not designed to precisely model curvilinear data such as those in Fig. 2, the interaction of nasal airflow and time was modeled using basis splines (B-splines; de Boor 1978:87–106), which use a piecewise polynomial function to simplify curvilinear data. The use of B-splines has the advantage of yielding coefficients that are more directly interpretable in terms of change over time than the orthogonal polynomials used in, for instance, growth-curve analysis. (In growth-curve analysis, individual coefficients are time-independent and represent the temporal curvature of the data only when combined.) Here, we use B-splines with three degrees of freedom to model the change in airflow over time using three different control points that dominate, very roughly, the first, middle, and final thirds of the vowel.[2] In our analyses, we label these Time 1, Time 2, and Time 3, but unlike analytic approaches in which time is binned and the dependent variable is averaged within each bin, these three temporal variables—in conjunction with the B-spline algorithm—describe the overall curvature of the data over time, and do not correspond to discrete time windows.

The dependent variable in the lmer model was nasal airflow (in ml/sec). Fixed effects were Time 1, Time 2, Time 3 (corresponding to the three B-spline coefficients), Voicing (voiced or voiceless coda in CVNC; reference level: voiceless), and their interactions. Participant and Word were included as random intercepts. Random slopes by Time and Voicing for Participant and Word were excluded, as the resulting models failed to converge. The production model's structure, using conventional lme4 syntax, is given in 1. The 'bs(time,df=3)' statement in 1 instructs the model to use B-splines to convert the original twenty-five-point measurements into curves, with three degrees of freedom (df=3), producing the three Time variables as described above.

(1) nas_flow_ml ~ bs(time,df=3)*voicing + (1|participant) + (1|word)

The predictions generated by the model are captured in Figure 3, which can be compared with the means of the actual data for the CVNC conditions in Fig. 2. The model captures both the overall shape of the curve and the subtle differences between the voiced and voiceless conditions. The changes in airflow over time, with the sharpest increases late in the vowel, are reflected by the effects of Time 1, Time 2, and Time 3 in Table 1,

[2] Three-degree-of-freedom splines were used as they provided a very good fit to both the airflow and perception data. Four- and five-degree-of-freedom B-splines were tested as well, but made only minor changes to curvature and did not change the main statistical outcomes.

which provides the model coefficients, along with an estimate of variance, for effects and interactions that reached significance. Voicing interacted with Time throughout the vowel, yet the overlapping 95% confidence intervals for the modeled curves in Fig. 3 indicate that the airflow differences between the voiced and voiceless contexts were in fact small.

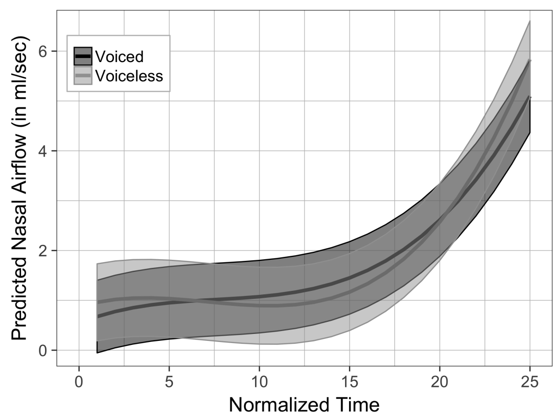

FIGURE 3. Model predictions of nasal airflow, over time, in voiced /CVnd/ (dark curve) and voiceless /CVnt/ (light) contexts. Upper and lower lines in each curve: 95% confidence intervals.

PREDICTOR	β	*SE*(β)	*t*	*p*
(intercept)	0.96	0.39	2.43	0.02
Time 1	0.59	0.07	8.82	< 0.001
Time 2	−2.21	0.05	−48.53	< 0.001
Time 3	4.87	0.03	139.36	< 0.001
Time 1 * Voicing	0.32	0.09	3.54	< 0.001
Time 2 * Voicing	1.32	0.06	21.27	< 0.001
Time 3 * Voicing	−0.45	0.05	−9.52	< 0.001

TABLE 1. Results of the nasal airflow model, over time (coefficients 1–3), for CVNC productions.

Although greater airflow differences between voiced and voiceless contexts, particularly early in the vowel, might have been expected based on the literature (see §1), other previously reported voicing-determined patterns of nasalization are more clearly upheld. For example, nasal consonants have been shown to be especially short in prevoiceless contexts, with the extent of vowel nasalization trading off with nasal consonant duration (e.g. Raphael et al. 1975, Busà 2007, Beddor 2009). Similarly, in this study, nasal consonants were approximately 30% shorter in /CVnt/ than in /CVnd/ tokens. A linear regression performed on average nasal airflow over the first third of the vowel (i.e. over the region of onset of coarticulatory nasalization for most speakers) against nasal consonant duration showed a very weak but significant negative correlation ($r^2 = 0.07$, $p < 0.001$), indicating that, the shorter the nasal consonant, the more extensive early vocalic nasal airflow tends to be. The correlation was stronger when performed on only the voiceless contexts (/CVnt/; $r^2 = 0.14$, $p < 0.001$), which show more variable /n/ duration than the voiced contexts.

Finally, to reduce the number of dimensions required to describe the airflow differences across different speakers' productions, the combined CVNC and NVN data were subjected to a functional principal component analysis (fPCA), using the built-in prcomp() function in R. Unlike a classical PCA, which combines many distinct features into components, the fPCA used here examines changes in the functional time course of vocalic nasal airflow by taking average flow at each measured time point for each speaker as input features, yielding components corresponding to orthogonal changes in the temporal pattern of airflow seen across the speakers. The decision to include NVN productions was based on comparison of CVNC-only with combined CVNC and NVN fPCA models. The outcomes of the two models were highly similar and, anticipating the findings of §4, led to very similar results in relation to the perception data. Given that speaker-specific airflow patterns for CVNC closely resemble each individual's own NVN patterns, we chose to include all of the information relevant to characterizing each speaker's coarticulatory patterns. Because dividing the data by voicing before fPCA did not result in meaningfully different curves for each speaker (as voiced and voiceless productions did not differ substantially; see Fig. 2), fPCA was done on voiced and voiceless data combined, thereby minimizing model complexity.

Figure 4 gives the mean normalized flow signal (plus and minus one standard deviation) of the first two components, PC1 (left) and PC2 (right), which together account for 97% (88% and 9%, respectively) of the variance in the data. Speakers whose mean nasal airflow traces are closer to the dashed light gray curve in each plot have high values for that PC, and those closer to the dashed dark gray curve have low values. As indicated by these curves, PC1 captures speaker differences in the amount of nasal airflow—differences that are likely due primarily to speaker physiology and possibly general articulatory setting. That is, the fPCA here effectively normalizes airflow by accounting for the large across-speaker differences in absolute nasal flow (PC1), therefore allowing other components to capture temporally relevant differences. PC2 captures time-course differences in airflow: speakers with higher PC2 values have more plateau-like airflow curves (except for the final portion of the vowel), whereas lower PC2 speakers have more cline-like curves. The plateau pattern bears a strong resemblance to airflow curves that have been reported for French phonemically nasalized vowels (e.g. Cohn 1990:98, Delvaux et al. 2008). The plateau-like vs. cline-like difference is similar to the variable airflow patterns described by Cohn (1990:153) for English speakers. Importantly, speakers with higher PC2 values also have higher volume of nasal airflow early in the vowel, consistent with greater temporal extent and magnitude of nasalization.

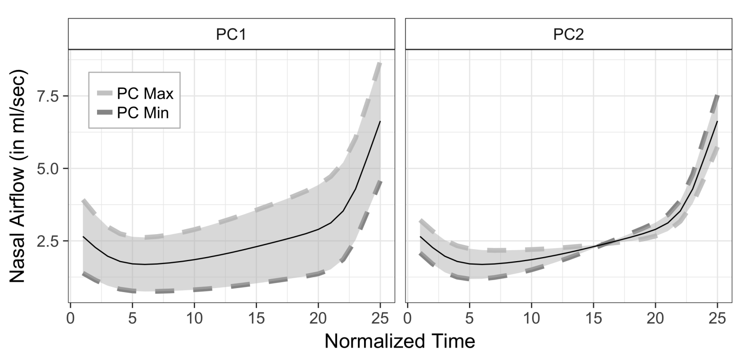

FIGURE 4. Mean normalized nasal airflow values (solid black line) for PC1 (left) and PC2 (right). Upper and lower dashed curves represent ±1 *SD*. The black line represents the average of the CVNC (voiced and voiceless) and NVN conditions from Fig. 2.

Of primary interest are the time-course differences—that is, early onset of plateaued nasal airflow compared to later onset of cline-like airflow—captured by PC2. PC2 provides a single value for each speaker that falls along a continuum ranging between the minimum and maximum values for this component. Speaker differentiation along this dimension emerges for both CVNC and NVN productions. These individual differences are illustrated in Figure 5, which gives the mean CVNC (left) and NVN (right) airflow curves for the ten most extreme (five highest and five lowest) speakers in terms of their PC2 values.

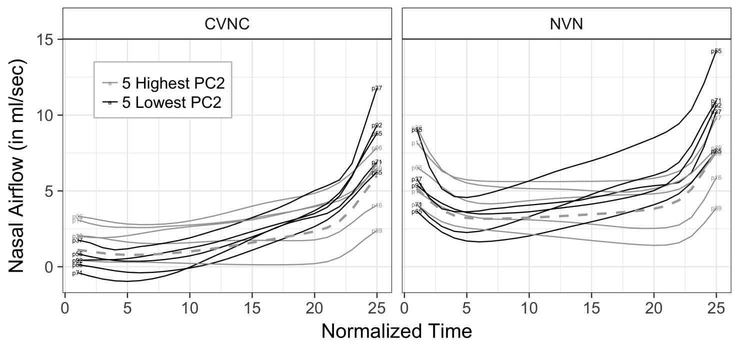

FIGURE 5. Nasal airflow curves over time, averaged across CVNC (left) and NVN (right) productions, for the five speakers with the highest (solid gray lines) and the five speakers with the lowest (solid black lines) PC2 values. Dashed line: mean overall flow for CVNC or NVN.

Our goal in this study is to link the time course of individuals' production of coarticulatory nasalization to the time course of their perception through statistical modeling. Although the results for coarticulatory nasal airflow generally show the expected context-specific contours, they exhibit as well systematic differences between individual speakers in the overall volume of nasal airflow, the temporal extent of coarticulatory airflow, and the plateau- or cline-like pattern of flow. The latter two differences, which reflect changes across the coarticulated vowel, are captured by PC2. Each participant's PC2 value therefore serves as the production measure that, in §4, is entered into the models of perception of coarticulatory nasalization.

3. PERCEPTION: THE PERCEPTUAL TIME COURSE OF COARTICULATORY NASALIZATION. The perception of coarticulatory nasalization is assessed using an eye-tracking design that replicates that of Beddor et al. 2013. We chose this design because that study provided clear evidence of systematic differences between participants in the time course of their use of the coarticulatory information.

In this experiment, participants hear an auditory CVC (e.g. *bed*) or CVNC (*bend*) word and see two visual images representing minimal-pair words (e.g. *bend-bed*); their task is to look at the image of the word they hear. The critical trials are those with auditory CVNC words, which were edited to create: (i) $C\tilde{V}_{early}NC$, with early onset of the coarticulatory information; (ii) $C\tilde{V}_{late}NC$, with later onset of that information; and (iii) $C\tilde{V}_{early}C$, with early coarticulatory information but no nasal consonant. Early and late onsets of vowel nasalization are characteristic of the variation that listeners hear in their everyday interactions. A nearly or completely absent nasal consonant, as in $C\tilde{V}_{early}C$, is also not uncommon in American English in voiceless contexts (*bent*, *scent*), but would be atypical in voiced contexts (*bend*, *send*). These manipulations were implemented to investigate the detailed timing patterns of participants' use of coarticulatory variation. (See §3.2 for specific predictions.)

3.1. METHODS.

PARTICIPANTS. The same forty-two undergraduates who participated in the production experiment also participated in the perception experiment. Among the additional participants who were recruited but eliminated (see §2.1) was one participant who failed to reach criterion on the critical auditory $C\tilde{V}_{early}NC$ and $C\tilde{V}_{late}NC$ trials. This individual reached only 45% target final fixations on the $C\tilde{V}_{early}NT$ (i.e. voiceless) trials and averaged only 70% target final fixations across all $C\tilde{V}NC$ trials. No other participant fell below 65% on any critical trial type, nor did any average less than 80% target final fixations across all $C\tilde{V}NC$ trials.

Stimuli. Stimuli were members of the five minimal quadruplets *bet-bed-bent-bend*, *let-lead-lent-lend*, *set-said-scent-send*, *wet-wed-went-wend*, and *watt-wad-want-wand*. Both auditory and visual stimuli were identical to those used in Beddor et al. 2013.

The auditory stimuli were modified versions of the twenty target words originally produced by a male phonetician who is a native speaker of Midwestern American English. The original recording included multiple repetitions of the twenty target words, from which two versions of each were selected on the basis of similarity to each other in vowel f0, duration, and formant frequencies. To control the time course of the coarticulatory information, these two versions were waveform edited in Praat. For each minimal CVNC-CVC word pair (e.g. *bent-bet*, *bend-bed*), the initial C and onset of V were taken from an original CVC token. To create the CVC stimulus (e.g. *bed*), this initial portion (e.g. be_{onset} from token 1) was spliced onto the offset V + final C of the second CVC token ($e_{offset}d$ from token 2). To create the corresponding CVNC stimulus (e.g. *bend*), the same initial portion (be_{onset} from token 1) was spliced onto the $V_{offset}NC$ of a CVNC token ([ɛ̃nd] from *bend*). In this approach, then, all stimuli are cross-spliced.

The two temporal degrees of vowel nasalization for CVNC stimuli were created by manipulating the proportion of the vowel taken from the original CVC and CVNC tokens. The proportion was 60% V from CVC and 40% $\tilde{V}$ from CVNC for late onset of nasalization ($C\tilde{V}_{late}NC$), and 20% V and 80% $\tilde{V}$ for early onset ($C\tilde{V}_{early}NC$). For many stimuli, these proportions were achieved by removing or duplicating a small number of pitch pulses of the original portions of the vowels. (For all nasal vowel portions, nasalization was clearly audible. Acoustic correlates were a decrease in waveform amplitude, and flattening and broadening of the F1 region of FFT spectra, relative to the oral portion.) Because vowels are longer in voiced than in voiceless contexts, the absolute duration of $\tilde{V}$ is longer in CVND than CVNT words, as indicated in Table 2.

	ORAL VOWEL	NASAL VOWEL	NASAL CONSONANT
$C\tilde{V}_{early}NT$	27	101	51
$C\tilde{V}_{late}NT$	79	51	51
$C\tilde{V}_{early}ND$	36	137	92
$C\tilde{V}_{late}ND$	99	75	92

Table 2. Average durations (in ms) of VN portions of CVNC stimuli.

The deleted-N $C\tilde{V}_{early}T$ and $C\tilde{V}_{early}D$ stimuli were designed to test listeners' sensitivity to context-specific patterns of gestural coordination (specifically, greater overlap of the lowered-velum gesture with the vowel constriction in voiceless contexts). These stimuli were created from stimuli with early onset of vowel nasalization by excising the nasal consonant, which was identified by its characteristic wave shape. These editing procedures resulted in four types of auditory stimuli: CVC, $C\tilde{V}_{early}NC$, $C\tilde{V}_{late}NC$, and $C\tilde{V}_{early}C$.

The target visual stimuli were twenty black-and-white line drawings corresponding to each of the twenty target words. (Additional images, drawn in the same style, were used in the practice trials.) Images were drawn by a professional artist, and each was sized to fit within a five-inch (72 dpi) square region of the computer screen.

In each test trial, participants heard one auditory stimulus and saw two visual stimuli. Paired images were always of minimal-pair words differing in the presence or absence of a nasal consonant or coda voicing, creating four types of visual pairings: CVT-CVD (*bet-bed*), CVT-CVNT (*bet-bent*), CVD-CVND (*bed-bend*), and CVNT-CVND (*bent-bend*). Each auditory stimulus was presented with each of the appropriate visual pairings except that, to reduce the size of the experiment, auditory $C\tilde{V}_{early}T/D$ occurred only with visual CVT/D-CVNT/D (and not with visual CVNT-CVND). Participants responded to 360 test trials. There were eighty visual CVT-CVD pairs, forty presented with auditory CVT and forty with auditory CVD (5 word pairs × 8 repetitions). There were also eighty visual CVNT-CVND: twenty (5 word pairs × 4 repetitions) each for auditory $C\tilde{V}_{early}NT$, $C\tilde{V}_{early}ND$, $C\tilde{V}_{late}NT$, and $C\tilde{V}_{late}ND$. There were 100 visual CVT-CVNT: twenty (5 pairs × 4 repetitions) each for auditory CVT, $C\tilde{V}_{late}NT$, $C\tilde{V}_{early}NT$, and $C\tilde{V}_{early}T$. The 100 visual CVD-CVND visual pairings are broken down as for the CVT-CVNT trials. Image position on the left or right of the screen was counterbalanced across trials.

PROCEDURE. Participants were tested individually in two sessions, with the test portion of each session consisting of half of the repetitions of each trial type and with the two testing sessions usually one week apart. In the first session, prior to testing, participants learned the labels for each of the twenty target images for the eye-tracking study and for the additional images for the production study. Participants first saw the randomly ordered images one at a time, with the label written below the image. To aid memorization, they read each label aloud to the experimenter and explained how the image related to the label. The second step of the familiarization procedure was identical to the familiarization described for the production experiment (§2.1). To ensure that familiarity was retained across sessions, this second step was also repeated at the beginning of the second perceptual testing session.

Eye movements were captured with a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research), using a 25 mm lens and sampling at 500 Hz. Participants were seated so that their eyes were between 550 and 650 mm from the camera and about 800 mm from the monitor. After familiarization but prior to testing, the experimenter performed a calibration procedure that, if necessary, was repeated until criterion was reached, typically on the dominant eye (unless that eye failed to track). During testing, auditory and visual stimuli were presented using SR Research Experiment Builder software; auditory stimuli were heard over AKG 271 Mk2 headphones.

In each test trial, participants saw a pair of images on the computer screen (e.g. *bed* and *bend*) and heard the recorded instruction 'Look at the pictures'. After two seconds, a fixation cross appeared in the center of the screen, as illustrated in Figure 6. Participants then heard 'Fixate cross. (pause) Now look at', followed by the target auditory stimulus (*bed* or *bend*). The fixation cross disappeared when the auditory target was played; the trial ended two seconds later. A five-minute break occurred halfway through each testing session. Before presentation of the test trials, participants responded to ten practice trials. No target stimulus was included in the practice trials.

DATA ANALYSIS AND PREDICTIONS. Participants' eye movements to the two images in each trial were monitored starting from the onset of the auditory stimulus and lasting 1000 ms. The computed measure was proportion target fixations over time, beginning at stimulus onset, for fifty 20 ms temporal bins. A fixation was counted as a target fixation if it fell within the target image's five-inch square region. Thus, a proportion of 0.50 for, say, the temporal bin 400–420 ms for auditory *bed* in a *bed-bend* trial means

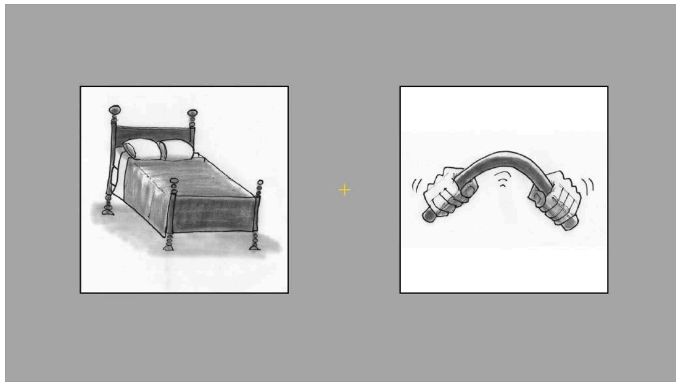

FIGURE 6. Screen shot for a *bed-bend* trial.

that 50% of those trials included a fixation on visual *bed* at some point during that 20 ms interval.

Although participants responded to auditory CVC stimuli, the main reason for their inclusion was to achieve a balanced design: visual pairings of CVC-CVNC require that participants hear both CVNC and CVC auditory stimuli. Here, we restrict analysis to responses to the auditory stimuli with vowel nasalization (i.e. $C\tilde{V}_{late}NC$, $C\tilde{V}_{early}NC$, and $C\tilde{V}_{early}C$) when the visual stimuli were CVC-CVNC—that is, to the stimuli for which we have predictions that we explore further in the subsequent production-perception comparisons in §4.

We hypothesize that, consistent with the findings of Beddor et al. 2013, participants will attend to the unfolding coarticulatory information and so will fixate the target image earlier and more often with auditory $C\tilde{V}_{early}NC$ than with $C\tilde{V}_{late}NC$ when the competitor image is CVC. This difference between early and late onset of vowel nasalization should be especially evident in the middle portion of the 0–1000 ms region of interest. Early vowel nasalization begins, on average, 106 ms after stimulus onset (102 ms for voiceless contexts and 111 ms for voiced); late nasalization begins 164 ms after stimulus onset (154 ms for voiceless and 174 ms for voiced). However, it takes time to program and launch an eye movement. Factoring in the typical estimate of a 200 ms programming delay (Dahan et al. 2001), if participants fixate the target image nearly as soon as they hear coarticulatory nasalization, the difference between responses to early and late nasalization stimuli should begin shortly after 300 ms from stimulus onset.

We hypothesize as well that, in general, listeners not only use available coarticulatory information in their communicative interactions, but also are sensitive to context-specific patterns of gestural overlap in the ambient language. As explained in §1, early vowel-nasalization onset and concomitantly short (or even absent) nasal consonants are more common in $VNC_{voiceless}$ than VNC_{voiced} sequences. Consequently, we predict that participants should be, over time, less likely to fixate the target CVNC image when the auditory stimulus lacks N ($C\tilde{V}_{early}C$) than when it includes N ($C\tilde{V}_{early}NC$), and that they should be especially unlikely to do so when prevoiced N is absent ($C\tilde{V}_{early}D$). For example, participants should be less willing to accept [bɛ̃d] than [bɛ̃t] as a CVNC word.

3.2. RESULTS. Our first goal is to establish that the predictions for listeners' overall responses, across participants, are upheld. Toward that end, the first two sections analyze the fixation patterns for the CVNC trials (Figs. 7 and 8) and the deleted-N trials (Fig. 9). The third section, relevant to this study's broader goal of establishing individual differences in the perceptual time course, presents data for two sample listeners to show that listeners differed in the time course of their fixation patterns (Fig. 10).

AUDITORY CṼNC TRIALS. This section reports the results for trials in which participants heard a CṼNC prompt and saw paired CVC and CVNC images. Figure 7 shows the pooled proportion fixations on the CVNC (i.e. target) image over time, beginning with stimulus onset, according to context and nasality. Because vowels are shorter in voiceless than in voiced contexts (see Table 2), both the disambiguating coarticulatory (Ṽ) and consonantal (N) cues occur earlier in voiceless trials. Consequently, there appears to be a higher proportion of target fixations in the voiceless (black curves) than in the voiced (gray) context beginning about 375 ms after stimulus onset. Stronger evidence that listeners track the unfolding coarticulatory information emerges in $C\tilde{V}_{early}NC$ (solid lines) compared to $C\tilde{V}_{late}NC$ (dashed) trials, which, within each voicing condition, differ only in the timing of coarticulatory nasalization. The expected higher proportion of target fixations in the early-nasalization condition also appears to emerge about 375 ms into the trial for the voiceless stimuli and slightly later for the voiced. Fixations on the target image begin well before onset of the nasal consonant, taking into account the time it takes to program and execute an eye movement, indicated by the 200 ms long arrows extending from the location of N onset.

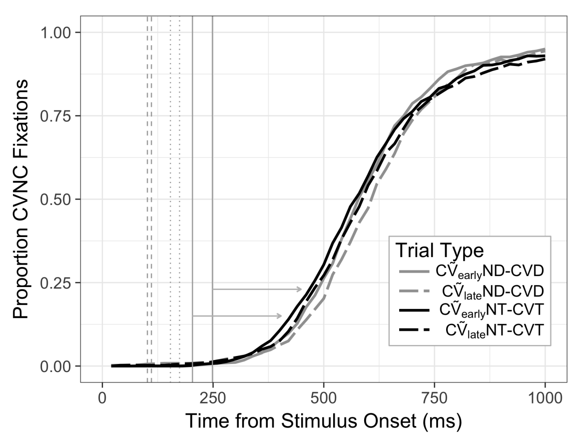

FIGURE 7. Pooled proportion target fixations, over time, on trials with auditory CṼNC according to degree of vowel nasalization (solid vs. dashed lines) and coda voicing (black vs. gray). Dashed vertical lines: early vowel nasalization onset; dotted: late vowel nasalization onset; solid: N onset. Left line of each type: voiceless context; right line: voiced context. Arrows indicate 200 ms eye-movement programming delay.

A generalized linear mixed-effects model was fit to the proportion CVNC fixations for trials with CṼNC prompts and CVC-CVNC images. Prior to analysis, data from 0–200 ms following stimulus onset were excluded because the eye-movement programming delay means that any target fixations during this period were likely not in response to the stimulus. In the model, the dependent variable was target fixations, with fixed effects of Time 1, Time 2, Time 3 (forty time points from 200 ms to 1000 ms after stimulus onset, again modeled with B-splines with three degrees of freedom), Nasalization (early, late; reference level: late), coda Voicing (voiced, voiceless; reference level: voiceless), and their interactions. Random intercepts were fitted for both Participant and Word, and random slopes were included for Word by Time and Nasalization. Random slopes were not included for Participant in order to keep this model's structure parallel with the production-perception model in §4.2, where we are explicitly investigating the presence of speaker-specific fixed effects.

Due to the binomial target-fixation data as well as the sheer size of the data set, conventional GLMER models were not able to converge using lme4. Consequently, all perception generalized linear mixed models (GLMMs) (§§3.2 and 4.2) were fit with a Bayesian approach, using the Markov chain Monte Carlo GLMM (MCMCglmm) package in R (Hadfield 2009). These models were fit using uninformative priors and default settings for mixed-effects binomial MCMCglmm models and were run using 150,000 sample Markov chains.[3] Model convergence was assessed using the Gelman-Rubin criterion (Brooks & Gelman 1998), where each model is run using multiple parallel chains, and the variances for each parameter are compared both within and among these chains, resulting in a potential scale reduction factor (PSRF) statistic. A fully converged model should show nearly identical variances and estimates across chains, with a PSRF less than 1.1 indicating that all chains, despite differences in the random sampling, have converged on essentially the same parameter estimates. Using conventional lme4 syntax for convenience, this model's structure is as in 2, where 'IsCVNCFixation' is a binary variable indicating whether the fixation has a CVNC target for this trial/time point. (The actual MCMCglmm command used to generate this and other models is provided in the supplementary materials, which can be accessed online at http://muse.jhu.edu/resolve/56.)

(2) IsCVNCFixation ~ bs(time,df=3)*voicing*nasalization + (1|participant) + (1+bs(time,df=3)+nasalization|word)

The predictions of the model are summarized in Figure 8, which can be compared with participants' actual fixation patterns in Fig. 7. As would be expected, the model predicts that fixations on the target image increase as time from stimulus onset increases. Also as expected, a higher proportion of target looks with $C\tilde{V}_{early}NC$ prompts (solid lines) than with $C\tilde{V}_{late}NC$ (dashed) is predicted. This $\tilde{V}_{early}$-$\tilde{V}_{late}$ difference in predicted responses emerges nearly as soon as the disambiguating coarticulatory information becomes available (see Fig. 7). It is also in this same region, which is dominated by the second B-spline control point, or Time 2, that the influence of voicing is evident: here the model predicts a higher proportion of target looks for voiceless C$\tilde{V}$NT prompts (black lines) than for voiced C$\tilde{V}$ND (gray), a difference that is due to the earlier occurrence of disambiguating information in the voiceless context.

Table 3 gives the results for those fixed factors and interactions that reached significance in the model. The table reports, for each coefficient, posterior means (i.e. the model's estimate for that coefficient). The statistical significance of each coefficient was assessed using model-derived 95% intervals of highest posterior density (HPD) (equivalent to 95% confidence intervals in other models), where an interval excluding zero represents a significant effect at the $p < 0.05$ level or lower. p_{MCMC}, which reflects this assessment in a more familiar manner, is also reported. All parameters in all reported models in this article show a PSRF of less than 1.05, indicating a fully converged model (Gelman & Shirley 2011).

[3] The robustness of these findings to autocorrelation (e.g. Baayen et al. 2018) was checked in two ways: first, by reanalyzing the perception data at coarser levels of temporal granularity (twenty rather than fifty bins) to confirm that the relevant effects persisted, and second, by excluding every other time point prior to entry into the model, then checking for the mathematically expected increase in modeled standard error ($\sim\sqrt{2}$). Both tests point to a minimal effect of autocorrelation in these models.

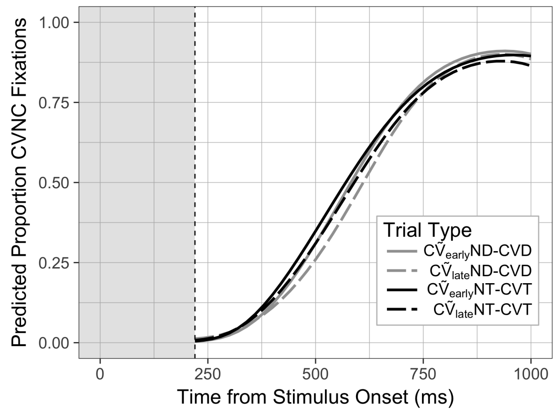

FIGURE 8. Model predictions of proportion target fixations to auditory CṼNC trials according to vowel nasality and coda voicing. (The shaded 0–200 ms region was excluded from the model.)

PREDICTOR	β	95% HPD INTERVAL		p_{MCMC}
(intercept)	−6.35	−7.22	−5.48	< 0.003
Time 1	4.33	1.84	7.34	< 0.003
Time 2	10.31	9.14	11.51	< 0.003
Time 3	9.38	8.37	10.44	< 0.003
Time 1 * Nasalization (Early)	1.81	0.91	2.88	0.006
Time 3 * Nasalization (Early)	1.03	0.54	1.67	0.006
Time 2 * Voicing (Voiced) * Nas. (Early)	0.82	0.20	1.40	0.01

TABLE 3. Results of the auditory CṼNC perception model.

AUDITORY DELETED-N TRIALS. Although participants' earlier looks to the CVNC image in response to $C\tilde{V}_{early}NC$ compared to $C\tilde{V}_{late}NC$ prompts demonstrate their overall attention to the coarticulatory information, trials with $C\tilde{V}_{early}C$ prompts test whether that information alone is sufficient for participants to sustain CVNC fixations. Participants' pooled responses in the left panel of Figure 9 show that those early fixations are, on average, sustained. Unsurprisingly, though, CVNC fixations in the absence of an acoustic nasal consonant do not begin to approach the greater than 90% fixations for $C\tilde{V}_{early}NC$ (Fig. 7), especially in the voiced condition.

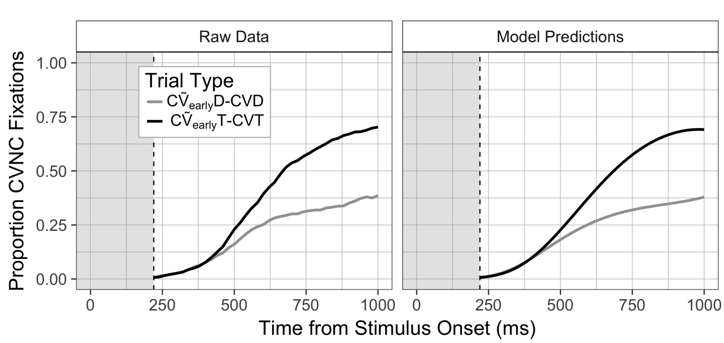

FIGURE 9. Participants' proportion target fixations (left panel) and model predictions (right) for trials with auditory $C\tilde{V}_{early}C$ according to coda voicing.

An MCMCglmm was fit to the proportion CVNC fixations for trials with $C\tilde{V}_{early}C$ prompts. This model was identical to that for the CṼNC auditory prompts except that it contained no fixed effect nor by-Word random slope for Nasalization (because all deleted-N stimuli have early onset of nasalization). The right panel in Fig. 9 gives the model predictions, and Table 4 gives the results for the fixed effects and interactions found to be significant in the model. The model predicts, as expected, more fixations with auditory voiceless $C\tilde{V}_{early}T$ (black curve) than with auditory voiced $C\tilde{V}_{early}D$ (gray). This voicing difference emerges near the point of N excision, which occurs 406 and 438 ms after stimulus onset (including a 200 ms programming delay; see Fig. 7) for voiceless and voiced contexts, respectively.

PREDICTOR	β	95% HPD INTERVAL		p_{MCMC}
(intercept)	−6.16	−6.79	−5.54	< 0.003
Time 1	4.94	3.29	6.95	< 0.003
Time 2	7.31	6.34	8.58	< 0.003
Time 3	7.41	6.38	8.28	< 0.003
Time 2 * Voicing (Voiced)	−2.46	−4.00	−0.99	< 0.003
Time 3 * Voicing (Voiced)	−2.09	−3.52	−0.79	0.007

TABLE 4. Results of the auditory $C\tilde{V}_{early}C$ perception model.

In summary, participants' eye movements as they listened to auditory prompts with coarticulatory vowel nasalization show that, overall, participants anticipate an upcoming nasal consonant shortly after the information about velum lowering becomes available. This outcome is most evident in responses to stimuli that differ in the timing of the onset of nasal coarticulation: the earlier the coarticulatory information, the faster and more accurate the fixations on the target image (Figs. 7 and 8). Participants' eye movements in response to auditory prompts with coarticulatory nasalization but no nasal consonant, however, show that listeners are responding to the information for velum lowering in context-dependent ways. As a group, participants weight more heavily the coarticulatory information in the voiceless context (Fig. 9), that is, in the context in which that information can be especially important in the Midwestern American English dialect spoken by many of these participants.

INDIVIDUAL DIFFERENCES. Just as these participants' production of coarticulatory nasalization varies across individuals, so too does the time course of their perception. Individual perceptual differences, and their relation to production, are systematically examined in §4. Here we briefly illustrate the nature of these differences for two participants, one from each group of the speakers with the highest and lowest PC2 values (that is, speakers with earlier and later onset of coarticulatory nasalization; Fig. 5). Figure 10 gives the responses of participants P17 (left; with a high PC2 value) and P55 (right; low PC2) to auditory $C\tilde{V}$(N)C prompts with a CVC competitor image. Although both participants fixated the target image by the end of trials in which the nasal consonant was present (solid lines), close scrutiny shows three differences: (i) P17 fixated the target image more quickly, reaching 50% target fixations about 100 ms before P55; (ii) P17's responses show greater differences between stimuli with early vs. late onset of nasalization (corresponding black vs. gray solid lines); and (iii) for P17 but not P55, $C\tilde{V}D$ prompts with no nasal consonant (dashed lines with triangles) are sufficient to elicit systematic percepts of a CVND word. These differences are consistent with P17 finding coarticulatory nasalization more useful than P55 in making linguistic decisions, at least in this laboratory setting.

4. DOES INDIVIDUALS' PRODUCTION OF COARTICULATION PREDICT THEIR PERCEPTION?

4.1. MODEL STRUCTURE AND PREDICTIONS. Having delineated the aerodynamic and perceptual time course of nasalization for the same group of participants, we turn to the

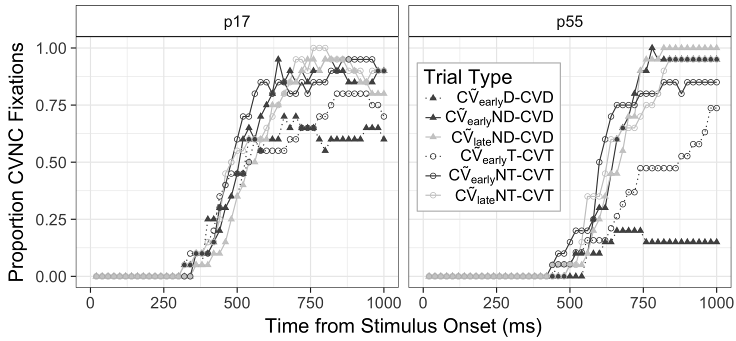

FIGURE 10. Proportion target fixations of two participants, P17 (left) and P55 (right), on trials with auditory CṼNC (solid lines) and CṼC (dashed) according to degree of vowel nasalization (black vs. gray lines) and coda voicing (triangles vs. circles). See text for discussion.

main question of this study: is there a relation between the production and perceptual use of this coarticulatory information at the level of the individual speaker-listener? In addressing this question, we asked whether a speaker's production is informative when predicting that individual's perceptual patterns. To this end, each speaker's production-based PC2 measure was incorporated into the perception data by labeling each perception trial with the corresponding speaker's overall PC2 value. These PC2 values were then entered into the two perception models presented in §3.2 as an additional predictor of perception, allowing it to interact with all other fixed effects.

We expect all participants, regardless of whether they produce early onset of coarticulatory vowel nasalization (as measured by PC2), to be accurate perceivers. For reasons offered in §1, however, we hypothesize that participants who produce earlier onset of the coarticulatory information will be, as listeners, more efficient or effective users of that information. This general prediction might be borne out in one or more ways in the perception models that incorporate production. Most importantly, we predict that participants who produce early onset of plateau-like nasalization—that is, participants with high PC2 values—will demonstrate earlier perceptual use of nasalization in auditory CṼNC trials by choosing the target image over its competitor more quickly than participants with low PC2 values. This hypothesized more efficient use of coarticulated nasality might also emerge in different patterns of responses to early compared to late onset of vowel nasalization. Specifically, participants who produce earlier onset of nasality might, if they are tracking the time-varying coarticulatory cues particularly closely, show a larger perceptual difference between the two degree-of-nasalization (early, late) conditions. Also possible is that a perceptual difference between early and late onset of nasalization might be temporally distributed, with participants with high PC2 values exhibiting a difference at an earlier point in time (within the 1000 ms period of interest) than those with low PC2 values.

For auditory stimuli in which the nasal consonant has been deleted (CṼC), we again predict that the perceptual time course will exhibit earlier fixations on the target CVNC image for participants who produce early onset of nasality. Moreover, recall that, overall, listeners were significantly less likely to look at the target CVNC image in voiced CṼD than in voiceless CṼT trials (Fig. 9), presumably because it is considerably less common for American English speakers to substantially shorten or omit a nasal consonant before voiced than before voiceless obstruents. Relatedly, a tentative prediction is that participants who produce late onset of nasality (i.e. have low PC2 values) might exhibit a particularly large perceptual voicing effect for CṼC trials, under the speculation that these individuals will be less likely to attend to nasalization in the (voiced) context in which that information is more routinely redundant.

4.2. Modeled results.

Auditory CṼNC trials. To assess whether individuals' production as measured by PC2 airflow values predicts their perception, a GLMM model using MCMCglmm was fit to the CṼNC (with CVC visual competitors) fixation data. The dependent variable was target fixations, with fixed effects of Time 1, Time 2, Time 3 (corresponding to the three B-splines), Nasalization (early, late; reference level: late), Voicing (reference level: voiceless), PC2, and their interactions. Participant and Word were included as random intercepts. Random slopes were included for Word by Time and Nasalization. Because this model explicitly seeks to investigate differences among speakers over time, random slopes for Participant were not included, as this would split the effect under study into different layers of the model, given that differences in fixation over time are being explicitly tested by the fixed effect of PC2.[4] (By comparison, random intercepts for Participant test overall differences in fixation proportion.) As in the previous models, this model used uninformative priors and 150,000 sample chains. Again using conventional lme4 syntax for convenience, 3 gives this model's structure, where PC2 corresponds to each participant's PC2 value.

(3) IsCVNCFixation ~ bs(time,df=3)*voicing*nasalization*pc2 + (1|participant) + (1+bs(time,df=3)+nasalization|word)[5]

Figures 11 and 13 summarize the model predictions. In the interest of displaying optimally representative values, these figures present model predictions for the mean value of PC2, as well as values corresponding to 90% and 10% of the overall PC2 range attested among the speakers.

Our main expectation is that the measure of production, PC2, will interact with perception, with participants with high PC2 values having earlier CVNC fixations than those with low PC2. This expectation is robustly borne out in the model. Figure 11 shows that the model predicts the expected effects of PC2 on proportion CVNC fixations in all nasality and voicing conditions. The expected effect of production disappears by about 700 ms after stimulus onset (or earlier, in the voiceless context). Given that the longest auditory stimulus was 475 ms long, this timing indicates that the perceptual difference between participants with high and low PC2 values continues throughout the auditory prompt (adding 200 ms for programming delay). In three conditions (all but 'early voiceless'), the fixation patterns reverse at about 800 ms after stimulus onset such that participants who produce early onset of nasality are predicted to decrease their CVNC fixations toward the end of the trial. This outcome is most likely due to these participants having earlier looks to the visual image and beginning to look away in later time windows (and accounts for the four-way interaction involving Time 3 in Table 5). (Trude & Brown-Schmidt 2012 also reported a late reversal in a

[4] A model run with full random slopes by Participant still showed a similar and significant PC2 effect, but with reduced magnitude, reflecting this relationship between fixed and random effects.

[5] In order to determine whether a simpler model would adequately fit the data, the production/perception models were also run with Voicing excluded (combining voiced and voiceless data into one larger model). Although the PC2 effect remained present in these simpler models, model comparison using DIC (deviance information criterion, a Bayesian version of AIC; cf. Gelman et al. 2014) indicated that the Voicing-included models provided better fit, showing reductions in DIC > 25 for models that include Voicing.

condition in which participants looked particularly early to the target image and offered a similar interpretation.) That these modeled patterns accurately capture the time course of the actual data can be seen by comparing Fig. 11 with Figure 12. Figure 12 gives the actual mean proportion fixations over time for the ten participants with the highest PC2 values and the ten participants with the lowest PC2 values, and it exhibits the patterns just described.

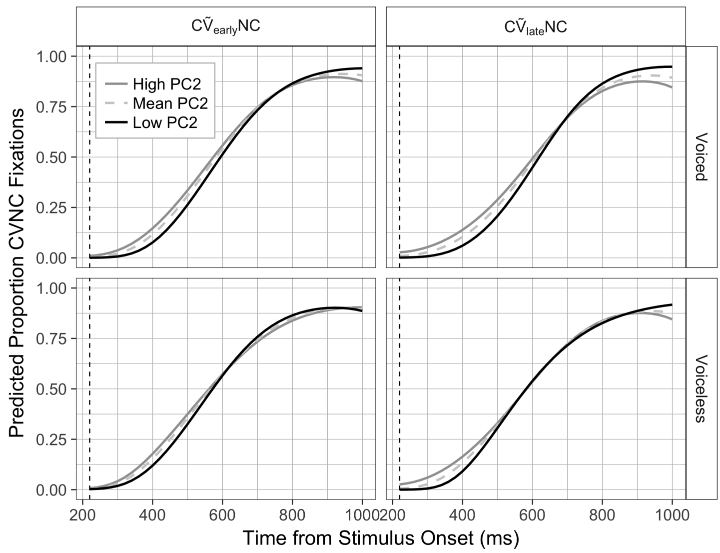

FIGURE 11. Model predictions of proportion target fixations to auditory CṼND (top) and CṼNT (bottom) trials in two nasality conditions according to participants' airflow PC2. Dashed lines: mean PC2 value. High (solid gray) and low (solid black) PC2 correspond to 90% and 10%, respectively, of overall PC2 range.

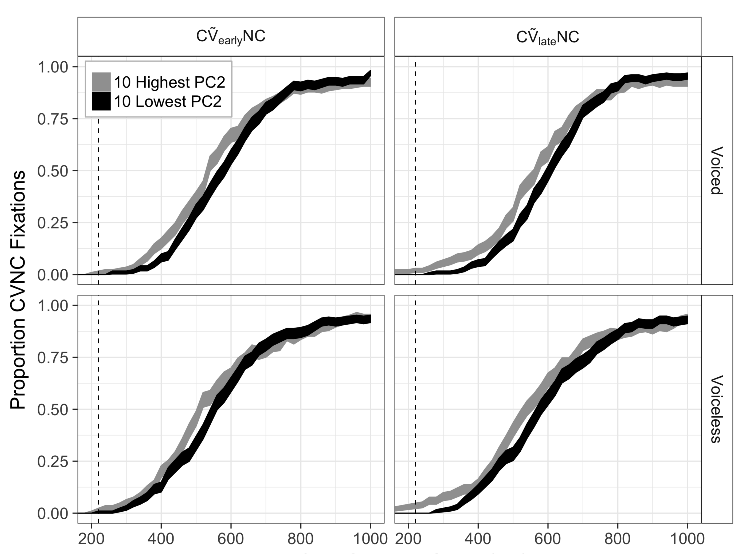

FIGURE 12. Mean proportion target fixations to auditory CṼNC trials of participants with ten highest (gray curves) and ten lowest (black) PC2 values. Line width represents standard error.

We also predicted (see §4.1) that PC2 should interact with Nasalization, with participants with high PC2 values showing an especially large perceptual difference between C$\tilde{V}_{early}$NC and C$\tilde{V}_{late}$NC trials. And we suggested that Time might also interact with PC2 and Nasalization such that high PC2 participants would exhibit a $\tilde{V}_{early}$-$\tilde{V}_{late}$ Nasalization difference at an earlier point in the trials than low PC2 participants. These predictions, which were upheld, are elucidated by inspecting both model fixation predictions and model results. Table 5 gives the model results for significant interactions involving PC2, without repeating the main effects of Time previously reported in Table 3. The most important takeaway from Table 5 is that PC2 participates in many interactions, as is expected if the perceptual use of coarticulation mirrors its production. These interactions align with our predictions except that the PC2 by Nasalization by Time interactions were restricted to the voiceless context. Although this effect is apparent in close comparison of the right and left lower panels of Fig. 11, for ease of reference Figure 13 reconfigures model predictions according to nasality. Figure 13 reveals that model predictions for high PC2 values show an earlier effect of early vs. late vowel nasality (400–650 ms) than do those for low PC2 values (600–900 ms after stimulus onset), consistent with participants who produce early coarticulatory nasalization using that information especially quickly in perception.

PREDICTOR	β	95% HPD INTERVAL		p_{MCMC}
(intercept)	−6.92	−7.74	−6.05	< 0.003
PC2	0.74	0.49	0.98	< 0.003
Time 1 * PC2	−1.28	−1.63	−0.95	< 0.003
Time 2 * PC2	−0.38	−0.55	−0.25	< 0.003
Time 3 * PC2	−0.88	−1.10	−0.69	< 0.003
Nasalization (Early) * PC2	−0.55	−0.76	−0.31	< 0.003
Time 1 * Nas. (Early) * PC2	1.21	0.74	1.73	< 0.003
Time 3 * Nas. (Early) * PC2	0.73	0.44	1.01	< 0.003
Time 1 * PC2 * Voicing (Voiced)	0.73	0.19	1.16	0.007
Nas. (Early) * PC2 * Voicing (Voiced)	0.48	0.18	0.83	< 0.003
Time 1 * Nas. (Early) * PC2 * Voicing (Voiced)	−1.18	−1.89	−0.51	< 0.003
Time 3 * Nas. (Early) * PC2 * Voicing (Voiced)	−0.59	−0.98	−0.16	0.007

TABLE 5. Results of the combined PC2 production/C$\tilde{V}$NC perception model. Only the significant interactions involving PC2 are reported.

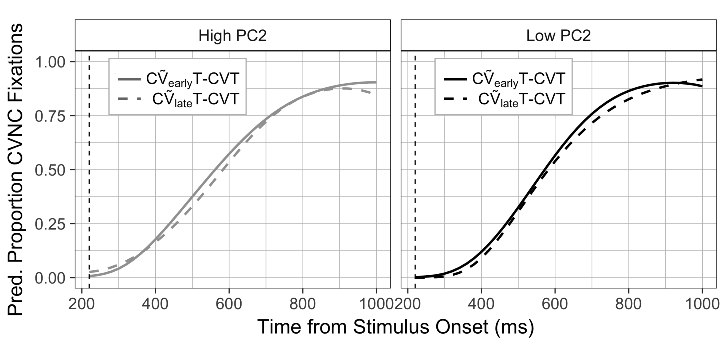

FIGURE 13. Model predictions for proportion target fixations to auditory C$\tilde{V}$NT (i.e. voiceless context) trials according to high and low PC2 values and vowel nasality.

Overall, although the differences in the perceptual patterns of participants who produce early compared to those who produce later onset of coarticulatory nasalization are small, the model of the auditory C$\tilde{V}$NC trials shows that, across conditions, participants' production as measured by PC2 airflow values contributes to predicting their perception. These differences are evident as well in the responses of individual speaker-listeners, as shown in Figure 14, which gives responses to CṼNC prompts for the ten participants (with five highest and five lowest PC2 values) whose production data were shown in Fig. 5. Although one participant with low PC2 values, P65, has unexpectedly early target fixations, it is otherwise the high PC2 speakers, with earlier, plateau-like nasal airflow patterns, who show earlier fixations, and low PC2 speakers, with later, cline-like articulations, who cluster together with later fixation times.

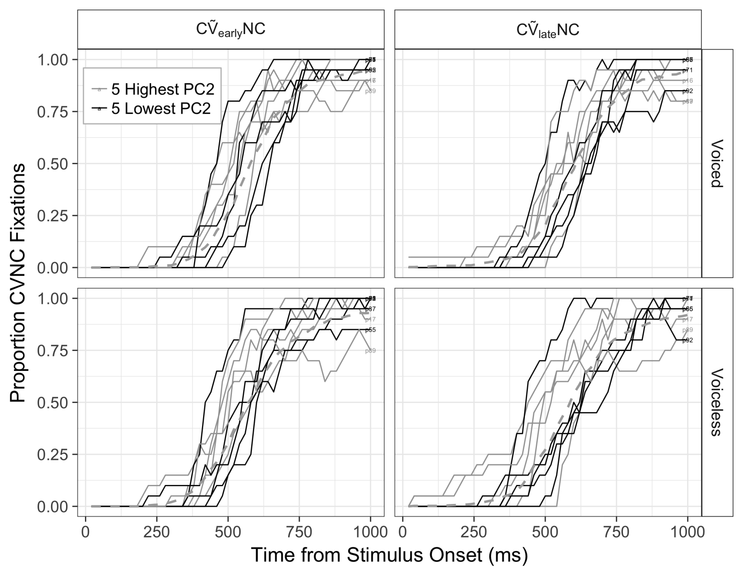

FIGURE 14. Proportion target fixations to auditory CṼNC trials of participants with five highest PC2 (gray) and five lowest (black) PC2 values.

AUDITORY DELETED-N TRIALS. An MCMCglmm with the same structure and settings, but without Nasalization as a fixed effect, was fit to the proportion CVNC fixations for trials with CṼ$_{early}$C prompts. The dependent variable was target fixations, and fixed effects were Time 1, Time 2, and Time 3, Voicing (reference level: voiceless), PC2, and their interactions, along with random intercepts for Word and Participant and random slopes for Word by Time. The model's structure is given in 4.

(4) IsCVNCFixation ~ bs(time,df=3)*voicing*pc2 + (1|participant) +
(1+bs(time,df=3) |word)

Participants who produce early coarticulatory nasalization are again expected to fixate CVNC images more quickly, with more looks to the target in the earlier region of the time course. These expectations are confirmed in the model's predictions in Figure 15 (left panel), which shows that PC2 impacts perception primarily during the early part of the fixation curve. The model predictions for the voiceless context (Fig. 15, lower left) also seem to predict a strong reversal in the later portion of the trial, which is supported by the model results in Table 6 (which again reports only significant effects involving PC2). Specifically, we see that PC2 interacts with Time throughout the vowel. The late reversal predicted by the model is, however, not apparent in participants' actual fixations: the mean fixations for participants with the ten highest and ten lowest PC2 values, given in the middle panel of Fig. 15, have overlapping distributions in the 700–

1000 ms region. We attribute the discrepancy between the modeled and actual data to participants' highly variable responses especially later in the $C\tilde{V}_{early}T$ trials.

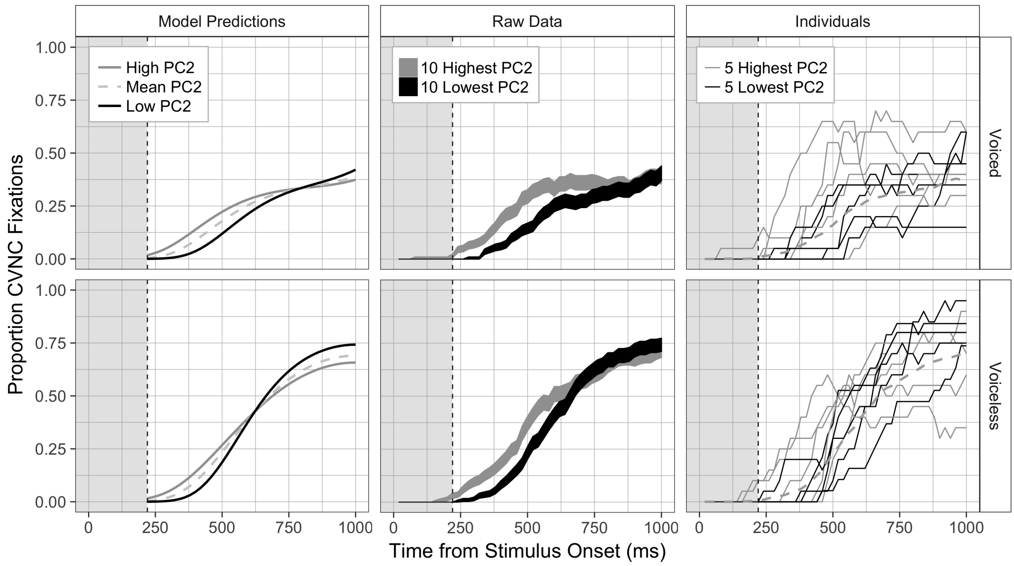

FIGURE 15. Proportion CVNC fixations to auditory $C\tilde{V}_{early}C$. Left: model predictions according to participants' airflow PC2. Middle: mean actual fixations of participants with ten highest and ten lowest PC2 values. Right: fixations of individual participants with five highest and five lowest PC2.

PREDICTOR	β	95% HPD INTERVAL		p_{MCMC}
(intercept)	−6.74	−7.45	−6.19	< 0.003
PC2	0.59	0.36	0.83	< 0.003
Time 1 * PC2	−0.66	−1.00	−0.33	< 0.003
Time 2 * PC2	−0.64	−0.77	−0.50	< 0.003
Time 3 * PC2	−0.67	−0.92	−0.52	< 0.003

TABLE 6. Results of the combined PC2 production/$C\tilde{V}_{early}C$ perception model. Only the significant effects involving PC2 are reported.

We suggested (§4.1) that, because vowel nasalization serves as critical information more often in $VNC_{voiceless}$ than in VNC_{voiced} contexts (e.g. differentiating *bet-bent* but more rarely *bed-bend*), participants who produce less coarticulatory nasalization might be especially unlikely to attend to this information in voiced contexts. The data, though, provide no evidence of a larger voicing difference, toward the end of the $C\tilde{V}_{early}C$ trials, for participants with low compared to high PC2 values. The group patterns are well illustrated by the $C\tilde{V}_{early}C$ perception results for the same ten participants (five highest and five lowest PC2 values) whose production (Fig. 5) and $C\tilde{V}NC$ perception (Fig. 14) data have already been shown. The right panel of Fig. 15 shows that high PC2 participants generally have earlier fixations on the CVNC (and especially the CVND) image but that, by the end of the trial, they are no more likely than low PC2 participants to fixate that image.

5. DISCUSSION.

5.1. SUMMARY. This study investigated the relation between individuals' articulatory and perceptual repertoires for coarticulatory nasalization. Grounded in the view that there are shared forms across the modalities of speaking and listening, and that these shared forms are complexly related at the level of individual language users, we hypothesized that individuals who more consistently and extensively produce coarticulatory nasalization in their own speech will attend to this information especially closely in perception. Speakers' patterns of nasal airflow revealed that, although all speakers produced coarticulatory nasal airflow during the vowel preceding a nasal consonant, the time course of the airflow patterns differed measurably and reliably across individuals. These individual differences were modeled in a functional principal component analysis whose second component, PC2, captured speaker differences in both the volume of nasal airflow early in the vowel and the changing slope of the airflow functions over time. Speakers whose nasal airflow patterns were associated with higher PC2 values produced relatively heavy coarticulatory flow early in the vowel and had plateau-like airflow functions (until shortly before the nasal consonant), whereas those whose patterns were associated with lower PC2 values had later onset of coarticulation and more cline-like curves (Figs. 4 and 5).

The time course of these same participants' perception of coarticulatory nasalization, as that information became available in real time, was measured in an eye-tracking task. As a group, listeners closely attended to the coarticulatory information: the earlier the coarticulatory cues for an upcoming nasal consonant, the earlier listeners look to target images representing CVNC rather than CVC words (Figs. 7 and 8). As a group, listeners also demonstrated that nasalization was especially informative in the voiceless context—that is, in the context in which the coarticulatory property tends, in the ambient speech of the community, to be particularly robust and the trigger of the coarticulation (the nasal consonant) is more variably realized (Fig. 9).

Statistical models that tested whether individual participants differed in their perceptual processing of coarticulatory nasalization, and whether individuals' perceptual patterns are partially predicted by their production of that information, provide support for the hypothesized relation between perception and production. Participants who produced heavier coarticulatory airflow early in the vowel (high PC2 values) looked earlier to, and had more fixations over time on, the target CVNC images in response to auditory prompts with vowel nasalization (Figs. 11 and 14). Consistent with these participants more closely tracking the coarticulatory information, they also exhibited (slightly) larger perceptual differences between the (early, late) conditions that differed in the temporal extent of vowel nasalization. Participants' perceptual time course differed as well in that the effect of the temporal manipulation of coarticulatory nasalization emerged relatively early in eye-tracking trials for participants with high PC2 values and only later for participants with lower values (at least in voiceless contexts; Fig. 13). However, not all expected links between production and perception of coarticulation were realized in the data. In particular, although participants differed in whether vowel nasalization alone, without an accompanying nasal consonant, was sufficient to systematically elicit CVNC lexical decisions (Fig. 15, rightmost panel), this perceptual difference did not depend on participants' production patterns. Put another way, these data do not provide evidence that individuals' production predicts their FINAL decision about the word they heard. This may be, in part, because the main perceptual condition in which final decisions substantially differed was also the only one in which the auditory stimulus, voiced CṼD, would rarely occur in natural speech, possibly leading some participants to use—in this laboratory setting—cues that they might otherwise find less informative (i.e. in settings where other information would be available). The data do, however, provide clear evidence that the time course of individuals' production of coarticulatory nasalization predicts, to some degree, the time course of their perception of that information.

5.2. The relation between individuals' production and perception. We consider first how to interpret the perceptual findings in light of theoretical approaches to speech perception that hold that, due to its lawful nature, coarticulatory variation informs perception (e.g. Strange et al. 1983, Whalen 1984, Fowler 2006; see §1). We share this perspective and, indeed, the related perspective that this informativeness should hold, to some extent, for all perceivers. In our view, the current perceptual results provide further support for this approach. Regardless of whether participants produced early or later onset of coarticulatory nasalization, they showed sensitivity to that information in the eye-tracking task. Perhaps the clearest evidence of this sensitivity is that producers of both early and later nasality (i.e. producers with high and low PC2 values) exhibited a difference between the two perceptual conditions that differed in the timing of the coarticulatory information. Where participants differed, as we have shown, is in the time course of their attention to coarticulation. This main finding effectively adds a temporal dimension to the individual differences in cue weighting that have been reported in the literature (e.g. Beddor 2009, Idemaru et al. 2012, Shultz et al. 2012, Schertz et al. 2015): here, some listeners attend more efficiently and, at least initially, more accurately than others to the evolving coarticulatory information.

These data also merit interpretation relative to previous studies, reviewed in §1, that have for the most part failed to establish a correlation between perception and production of coarticulation at the level of individual speaker-listeners (Grosvald 2009, Kataoka 2011, Grosvald & Corina 2012, Shultz et al. 2012, Schertz et al. 2015; though cf. Zellou 2017). A similarly negative outcome could quite possibly have been obtained in this study as well if instead of analyzing the time course of perception we had analyzed participants' final lexical choices, which were, as we have observed, not demonstrably tied to their production patterns. That the time-course analysis was successful in establishing a production-perception relation is, we believe, due to these data revealing the moment-by-moment usefulness of specific properties of the acoustic signal as they become available to a listener. The findings show that individuals who are the more efficient users of the coarticulatory cues in turn produce similarly informative (i.e. more heavily coarticulated) speech signals. Although it is somewhat tempting to suggest that these individuals do so because they find such signals useful to themselves as perceivers, we remain agnostic about the direction of this relation (e.g. perhaps routinely producing the strongly coarticulated signals more finely attunes an individual to that information).

5.3. Implications for sound change. A challenge for research on sound change, as aptly characterized by Stevens and Harrington, is to 'link the initiation of sound change within individual cognitive grammars with the diffusion of novel variants through the community' (2014:2). This study takes up one specific aspect of this challenge: to identify a mechanism, for perceptually motivated sound changes, through which a listener's percept—most relevantly, an innovative percept—is publicly manifested and hence has the potential to spread through the community. We investigated coarticulation because sound changes involving coarticulatory processes have become the classic example of changes in which the source of the phonetic variation is articulatory, but the impetus for the change involving those variants is arguably perceptual (e.g. Ohala 1981, 1993, Guion 1998, Harrington et al. 2008, Beddor 2009, Grosvald & Corina 2012, Yu 2013; though cf. Bybee 2012). This classic example and other claims of perceptually motivated changes rely, as discussed in §1, on listeners-turned-speakers producing variants that reflect their percepts. The main finding of a link between an individual's perception and production, with earlier and more efficient perceptual attention to coarticulation being tied to greater production of that information, offers empirical support for this assumption.

How might the perception-production relation established here help explain sound changes leading to distinctive vowel nasalization, whereby the coarticulatory trigger (N) is lost and the original coarticulatory effect (Ṽ) becomes the contrastive property? We refer to listeners who closely track coarticulatory nasalization, and attend to this information as soon as it becomes available to them, as 'innovative' listeners. This is not to say that all listeners who find coarticulatory information to be perceptually useful are innovative (indeed, we expect such processing to be the norm). Rather, in the context of describing incipient perceptually motivated sound change, we take listeners who fall at the more extreme, positive end of the continuum of attention to coarticulatory detail to be innovative. Some of these listeners are especially innovative in that not only do they readily anticipate a CVNC word on the basis of vowel nasalization, but they also do not require the nasal consonant to sustain the CVNC percept (see §4.2). The findings for participants with high airflow PC2 values indicate that the innovative listeners also tend to be innovative speakers in that they produce especially heavy anticipatory nasal coarticulation early in the vowel. It remains to be shown, then, how this situation—which is presumably one of stable variation in a speech community—could progress to one in which an emerging new norm increasingly has extensive vowel nasalization and decreasingly realizes the nasal consonant.

One possible account appeals to the trade-off, within stable variation, that has been reported between the temporal extent of vowel nasalization and the duration of a following nasal consonant due to variable alignment, in VNC sequences, of a relatively constant-sized lowered-velum gesture relative to oral vowel and stop constrictions (see §2.2 and Beddor 2009). This trading relation also holds for the current production data: in general, the earlier the onset of anticipatory nasal airflow during the vowel, the shorter the following nasal consonant (§2.2). Thus, innovative perception is linked, in production, to both more extensive vowel nasalization and shorter nasal consonants—that is, to nasal consonants that are, arguably, perceptually not highly salient for these individuals' interlocutors. These variants, if spread through the speech community, could, over time, potentially contribute to a change from ṼN to $\tilde{V}_N$ (i.e. shortened and only sporadically realized N) to Ṽ. Indeed, as we have discussed, $\tilde{V}_N$ is already the situation for phonological $VNC_{voiceless}$ sequences for some speakers of American English.

Figure 16 gives further evidence from the current production and perception data that the participants who show earlier attention to vowel nasality produce not only more extensively nasalized vowels but also shorter nasal consonants. Rather than taking our previous approach of evaluating the mean proportion CVNC fixations for the participants with the five highest and five lowest PC2 values, we instead calculated CVNC fixation curves for the participants with the five longest and five shortest average NASAL CONSONANT DURATIONS in voiceless contexts. The expected difference—that participants who produce shorter nasal consonants (and concomitantly earlier vowel nasalization) look earlier to the CVNC images—emerges only very early (up to about 400 ms after stimulus onset) in response to auditory $C\tilde{V}_{early}NC$ prompts (right panels) but extends through more of the perceptual trial in response to CṼC prompts (left).

That a trading relation between coarticulatory source and effect might contribute to change is consistent with the sound change literature. The synchronic tendency for earlier onset of vowel nasalization to cooccur with shorter nasal consonants in prevoiceless contexts, for example, is mirrored in the historical development of phonemic

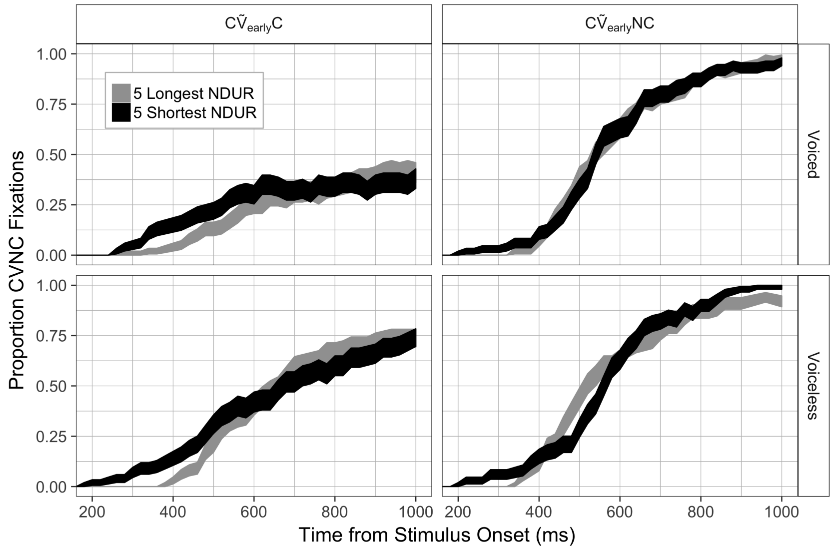

FIGURE 16. Proportion CVNC fixations to auditory $C\tilde{V}_{early}C$ (left) and $C\tilde{V}_{early}NC$ (right) trials for participants who produced the five longest and five shortest average N durations in voiceless contexts. Line width represents standard error.

vowel nasalization. Specifically, historical data from Romance (e.g. Hajek 1997:141–43, Sampson 1999:224, 274) and Indo-Aryan (Ruhlen 1978) languages indicate earlier development of vowel nasalization and nasal consonant loss in voiceless contexts. Three other coarticulatory trading relations that likely contribute to sound change are also relevant here. First, strikingly parallel to the realignment of the lowered-velum gesture that presumably contributes to phonemic nasalization is the realignment of an open-glottis gesture relative to stop closure in Western Andalusian Spanish. This realignment in /s/ + stop clusters yields a trading relation between preaspiration (e.g. [st], [ht]) and postaspiration (e.g. [t^h]) and is the source of change from unaspirated to aspirated stops in this variety (Parrell 2012). A second example is the relative weighting of VOT and f0 as information for voicing contrasts (see §1). These properties trade off, with longer positive VOTs for voiceless plosives in English, for example, cooccurring with lower f0 values (Shultz et al. 2012, Dmitrieva et al. 2015). For English, the VOT/f0 trade-off is presumably stable covariation. For Seoul Korean, however, which is undergoing tonogenesis, some findings point toward decreasing VOT differences between phonation types occurring in tandem with increasing f0 differences on following vowels (e.g. Kang 2014). Third, another trade-off implicated in (noninitial) voicing contrasts is that between the durations of stop closure and a preceding vowel: longer closures for voiceless stops cooccur with shorter vowels. Harrington et al. 2012 reported that a postvocalic voicing contrast is developing in the stop productions of younger speakers of the East Franconian dialect of German apparently due to an emerging trading relation between the source (in this case, a strongly released phonologically voiceless stop) and its effect (truncation of the preceding vowel). We view these trading relations as an important mechanism by which new production and perception norms may emerge from coarticulatory variation.

We are necessarily cautious in our interpretation of our results' implications for sound change. The 'innovative' listeners who rely especially heavily on vowel nasality early in the time course of perception are not the only individuals who, as speakers, produce early heavy coarticulatory nasalization. From the production side, not all producers of relatively late onset of coarticulation—that is, conservative speakers—are also conservative listeners. Indeed, for some of the coarticulatorily conservative speakers in this study, vowel nasalization alone (without a following nasal consonant) was sufficient for eliciting and sustaining looks to CVNC words (Fig. 15). Participants in this study do not fall neatly along an innovative-to-conservative perception-production continuum. Rather, our finding is that there is a small but significant and systematic tendency, demonstrated by statistical modeling, for perception and production of coarticulation to be linked. And, again, it is via this tendency that innovative—or conservative—perception can be manifested in production. Grosvald and Corina (2012) pointed out that, for listeners to contribute to sound changes involving coarticulation, a speech community as a whole would not need to exhibit a perception-production correlation; that is, it should be sufficient for SOME listeners to manifest their percepts in their coarticulated productions, and it is these quite possibly sporadic speaker-listeners who could initiate the change. Stevens and Harrington (2014) further suggested that, if indeed only very few individuals in a community have the requisite perception-production link, this might help explain why sound change is not rampant despite the phonetic conditions for sound change being ever-present in the ambient language. The current data, however, indicate that although perception of coarticulation is highly variable across participants in ways not explained by their coarticulated productions, production nonetheless partially predicts perception in expected ways.

We are also necessarily cautious in interpreting our results because these data do not address why particular individuals differ in their attention to, and production of, coarticulation. That individuals differ follows, we believe, from the multiple sources of information that are available for a given speech contrast and the typical status of coarticulatory information as redundant, subphonemic detail. However, in this study we lack participant information that might inform these individuals' specific (but linked) articulatory and perceptual repertoires. We expect that both linguistic experience and sociocognitive characteristics contribute not only to a given speaker-listener's patterns, but also to the strength of the link between these patterns, which varies across individuals. Although this speculation cannot be pursued further here, in ongoing research in our lab we are addressing some of these issues by extending investigation of the time course of perception and production to socially indexed patterns of coarticulatory variation.

6. CONCLUSION. Theories of phonetics and theories of the actuation of sound change share the goal of determining the relation between speech perception and production for individual speaker-listeners. This study of speaker-listeners' production of, and perceptual attention to, coarticulated speech demonstrates that listeners who are efficient users of the coarticulatory information and who track the coarticulatory information especially closely as the acoustic signal unfolds over time produce, as speakers, especially early onset of that same information. Other speaker-listeners, however, evidence a later time course in both perception and production. For theories of phonetics, this main finding suggests that the parity between the forms of speaking and listening is not limited to the requirement of sufficient similarity between these forms, but apparently extends to the production-perception relation within the individual language user. For theories of sound change, this outcome substantiates the long-standing assumption, especially of models of perceptually motivated change, that listeners-turned-speakers' productions reflect their perception. A speaker-listener's perception of coarticulated speech is made public through their productions.

REFERENCES

Baayen, R. Harald; Jacolien van Rij; Cecile de Cat; and Simon Wood. 2018. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *Mixed effects regression models in linguistics*, ed. by Dirk Speelman, Kris Heylen, and Dirk Geeraerts, 49–70. Cham: Springer.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40.177–89. DOI: 10.1016/j.wocn.2011.09.001.

Bates, Douglas; Martin Maechler; Ben Bolker; and Steven Walker. 2014. lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-6. Online: https://cran.r-project.org/web/packages/lme4/.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85.785–821. DOI: 10.1353/lan.0.0165.

Beddor, Patrice Speeter. 2012. Perception grammars and sound change. In Solé & Recasens, 37–55.

Beddor, Patrice Speeter; James D. Harnsberger; and Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30.591–627. DOI: 10.1006/jpho.2002.0177.

Beddor, Patrice Speeter, and Rena Arens Krakow. 1999. Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America* 106.2868–87. DOI: 10.1121/1.428111.

Beddor, Patrice Speeter; Kevin B. McGowan; Julie E. Boland; Andries W. Coetzee; and Anthony Brasher. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133.2350–66. DOI: 10.1121/1.4794366.

Boersma, Paul, and David Weenick. 2013. Praat: Doing phonetics by computer. Version 6.0.05. Online: http://www.praat.org/.

Boyce, Suzanne E. 1990. Coarticulatory organization for lip rounding in Turkish and English. *The Journal of the Acoustical Society of America* 88.2584–95. DOI: 10.1121/1.400349.

Brooks, Stephen P., and Andrew Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7.434–55. DOI: 10.1080/10618600.1998.10474787.

Busà, M. Grazia. 2007. Coarticulatory nasalization and phonological developments: Data from Italian and English nasal-fricative sequences. *Experimental approaches to phonology*, ed. by Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala, 155–74. Oxford: Oxford University Press.

Bybee, Joan. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. In Solé & Recasens, 211–34.

Coetzee, Andries W.; Patrice Speeter Beddor; Kerby Shedden; Will Styler; and Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66.185–216. DOI: 10.1016/j.wocn.2017.09.009.

Cohn, Abigail C. 1990. *Phonetic and phonological rules of nasalization*. (UCLA Working Papers in Phonetics 76.) Los Angeles: University of California, Los Angeles. Online: https://escholarship.org/uc/item/1xq3d5hr.

Dahan, Delphine; James S. Magnuson; Michael K. Tanenhaus; and Ellen M. Hogan. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16.507–34. DOI: 10.1080/01690960143000074.

de Boor, Carl. 1978. *A practical guide to splines*. (Applied mathematical sciences 27.) New York: Springer.

Delvaux, Véronique; Didier Demolin; Bernard Harmegnies; and Alain Soquet. 2008. The aerodynamics of nasalization in French. *Journal of Phonetics* 36.578–606. DOI: 10.1016/j.wocn.2008.02.002.

Diehl, Randy L.; Andrew J. Lotto; and Lori L. Holt. 2004. Speech perception. *Annual Review of Psychology* 55.149–79. DOI: 10.1146/annurev.psych.55.090902.142028.

Dmitrieva, Olga; Fernando Llanos; Amanda A. Shultz; and Alexander L. Francis. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49. 77–95. DOI: 10.1016/j.wocn.2014.12.005.

Flagg, Elissa J.; Janis E. Oram Cardy; and Timothy P. L. Roberts. 2006. MEG detects neural consequences of anomalous nasalization in vowel–consonant pairs. *Neuroscience Letters* 397.263–68. DOI: 10.1016/j.neulet.2005.12.034.

Fowler, Carol A. 2005. Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics* 33.199–213. DOI: 10.1016/j.wocn.2004.10.003.

Fowler, Carol A. 2006. Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics* 68.161–77. DOI: 10.3758/BF03193666.

Fowler, Carol A. 2007. Speech production. *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell, 489–501. Oxford: Oxford University Press.

Fowler, Carol A., and Julie M. Brown. 2000. Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics* 62.21–32. DOI: 10.3758/BF03212058.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press.

Gelman, Andrew; Jessica Hwang; and Aki Vehtari. 2014. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24.997–1016. DOI: 10.1007/s11222-013-9416-2.

Gelman, Andrew, and Kenneth Shirley. 2011. Inference from simulations and monitoring convergence. *Handbook of Markov chain Monte Carlo*, ed. by Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, 163–74. Boca Raton: CRC Press.

Grosvald, Michael. 2009. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics* 37.173–88. DOI: 10.1016/j.wocn.2009.01.002.

Grosvald, Michael, and David Corina. 2012. The production and perception of subphonemic vowel contrasts and the role of the listener in sound change. In Solé & Recasens, 77–100.

Guion, Susan Guignard. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55.18–52. DOI: 10.1159/000028423.

Hadfield, Jarrod D. 2009. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* 33(2). DOI: 10.18637/jss.v033.i02.

Hajek, John. 1997. *Universals of sound change in nasalization*. Boston: Blackwell.

Hajek, John. 2013. Vowel nasalization. *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath, Ch. 10. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: http://wals.info/chapter/10, accessed July 22, 2017.

Harrington, Jonathan; Felicitas Kleber; and Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123. 2825–35. DOI: 10.1121/1.2897042.

Harrington, Jonathan; Felicitas Kleber; and Ulrich Reubold. 2012. The production and perception of coarticulation in two types of sound changes in progress. *Speech planning and dynamics* (Speech production and perception), ed. by Susanne Fuchs, Melanie Weirich, Daniel Pape, and Pascal Perrier, 33–55. Frankfurt: Peter Lang.

Harrington, Jonathan; Felicitas Kleber; Ulrich Reubold; Florian Schiel; and Mary Stevens. 2019. The phonetic basis of the origin and spread of sound change. *The Routledge handbook of phonetics*, ed. by William F. Katz and Peter F. Assmann. New York: Routledge, to appear.

Idemaru, Kaori; Lori L. Holt; and Howard Seltman. 2012. Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America* 132.3950–64. DOI: 10.1121/1.4765076.

Iskarous, Khalil. 2005. Patterns of tongue movement. *Journal of Phonetics* 33.363–81. DOI: 10.1016/j.wocn.2004.09.001.

Kang, Yoonjung. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45.76–90. DOI: 10.1016/j.wocn.2014.03.005.

Kataoka, Reiko. 2011. *Phonetic and cognitive bases of sound change*. Berkeley: University of California, Berkeley dissertation. Online: https://escholarship.org/uc/item/3m20112c.

Kleber, Felicitas; Jonathan Harrington; and Ulrich Reubold. 2012. The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech* 55.383–405. DOI: 10.1177/0023830911422194.

Kong, Eun Jong, and Jan Edwards. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59.40–57. DOI: 10.1016/j.wocn.2016.08.006.

Krakow, Rena A. 1989. *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. New Haven, CT: Yale University dissertation.

Krakow, Rena A.; Patrice S. Beddor; Louis M. Goldstein; and Carol A. Fowler. 1988. Coarticulatory influences on the perceived height of nasal vowels. *The Journal of the Acoustical Society of America* 83.1146–58. DOI: 10.1121/1.396059.

Krakow, Rena A., and Marie K. Huffman. 1993. Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction. *Nasals, nasalization, and the velum*, ed. by Marie K. Huffman and Rena A. Krakow, 3–59. New York: Academic Press.

Kuznetsova, Alexandra; Per Bruun Brockhoff; and Rune Haubo Bojesen Christensen. 2015. lmerTest: Tests in linear mixed effects models. R package version 2(0). Online: https://cran.r-project.org/web/packages/lmerTest/.

Kwon, Harim. 2015. *Cue primacy and spontaneous imitation: Is imitation phonetic or phonological?* Ann Arbor: University of Michigan dissertation.

Liberman, Alvin M., and Doug H. Whalen. 2000. On the relation of speech to language. *Trends in Cognitive Sciences* 4.187–96. DOI: 10.1016/S1364-6613(00)01471-6.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

Lindblom, Björn; Susan Guion; Susan Hura; Seung-Jae Moon; and Raquel Willerman. 1995. Is sound change adaptive? *Rivista di Linguistica* 7.5–36.

Lubker, James, and Thomas Gay. 1982. Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *The Journal of the Acoustical Society of America* 71.437–48. DOI: 10.1121/1.387447.

Magen, Harriet S. 1997. The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25.187–205. DOI: 10.1006/jpho.1996.0041.

Malécot, André. 1960. Vowel nasality as a distinctive feature in American English. *Language* 36.222–29. DOI: 10.2307/410987.

Mann, Virginia A., and Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28.213–28. DOI: 10.3758/BF03204377.

Manuel, Sharon Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America* 88.1286–98. DOI: 10.1121/1.399705.

Martin, James G., and H. Timothy Bunnell. 1982. Perception of anticipatory coarticulation effects in vowel–stop consonant–vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance* 8.473–88. DOI: 10.1037/0096-1523.8.3.473.

Mitterer, Holger, and Mirjam Ernestus. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109.168–73. DOI: 10.1016/j.cognition.2008.08.002.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39.132–42. DOI: 10.1016/j.wocn.2010.12.007.

Noiray, Aude; Marie-Agnès Cathiard; Lucie Ménard; and Christian Abry. 2011. Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America* 129.340–49. DOI: 10.1121/1.3518452.

Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17(2).178–203.

Ohala, John J. 1993. Coarticulation and phonology. *Language and Speech* 36.155–70. DOI: 10.1177/002383099303600303.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119.2382–93. DOI: 10.1121/1.2178720.

Pardo, Jennifer S. 2012. Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language and Linguistics Compass* 6.753–67. DOI: 10.1002/lnc3.367.

Parrell, Benjamin. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics* 40.37–45. DOI: 10.1016/j.wocn.2011.08.004.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency effects and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137–57. Amsterdam: John Benjamins.

Pinget, Anne-France. 2015. *The actuation of sound change*. Utrecht: Utrecht University dissertation.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.r-project.org/.

Raphael, Lawrence J.; M. F. Dorman; Frances Freeman; and Charles Tobin. 1975. Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech, Language, and Hearing Research* 18.389–400. DOI: 10.1044/jshr.1803.389.

Ruhlen, Merritt. 1978. Nasal vowels. *Universals of human language: Phonology*, ed. by Joseph H. Greenberg, 203–41. Stanford, CA: Stanford University Press.

Sampson, Rodney. 1999. *Nasal vowel evolution in Romance*. Oxford: Oxford University Press.

Samuel, Arthur G. 2011. Speech perception. *Annual Review of Psychology* 62.49–72. DOI: 10.1146/annurev.psych.121208.131643.

Schertz, Jessamyn; Taehong Cho; Andrew Lotto; and Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52.183–204. DOI: 10.1016/j.wocn.2015.07.003.

Shosted, Ryan; Christopher Carignan; and Panying Rong. 2012. Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *The Journal of the Acoustical Society of America* 131.455–65. DOI: 10.1121/1.3665998.

Shultz, Amanda A.; Alexander L. Francis; and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America* 132.EL95–EL101. DOI: 10.1121/1.4736711.

Solé, Maria-Josep, and Daniel Recasens (eds.) 2012. *The initiation of sound change: Production, perception, and social factors*. Amsterdam: John Benjamins.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Mary, and Jonathan Harrington. 2014. The individual and the actuation of sound change. *Loquens* 1(1):e003. DOI: 10.3989/loquens.2014.003.

Strange, Winifred; James J. Jenkins; and Thomas L. Johnson. 1983. Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America* 74.695–705. DOI: 10.1121/1.389855.

Trude, Alison M., and Sarah Brown-Schmidt. 2012. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes* 27.979–1001. DOI: 10.1080/01690965.2011.597153.

Whalen, D. H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics* 35.49–64. DOI: 10.3758/BF03205924.

Whalen, D. H. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18.3–35.

Whalen, D. H.; Arthur S. Abramson; Leigh Lisker; and Maria Mody. 1993. *F*0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America* 93.2152–59. DOI: 10.1121/1.406678.

Yu, Alan C. L. 2010. Perceptual compensation is correlated with individuals' 'autistic' traits: Implications for models of sound change. *PLoS ONE* 5(8):e11950. DOI: 10.1371/journal.pone.0011950.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 201–27. Oxford: Oxford University Press.

Yu, Alan C. L. 2016. Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America* 139.1672–90. DOI: 10.1121/1.4944992.

Yu, Alan C. L., and Hyunjung Lee. 2014. The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *The Journal of the Acoustical Society of America* 136.382–88. DOI: 10.1121/1.4883380.

Zellou, Georgia. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61.13–29. DOI: 10.1016/j.wocn.2016.12.002.

[beddor@umich.edu]
[coetzee@umich.edu]
[wstyler@ucsd.edu]
[kbmcgowan@uky.edu]
[jeboland@umich.edu]

[Received 28 August 2017;
revision invited 19 January 2018;
revision received 29 May 2018;
accepted 21 June 2018]

PROJECT MUSE

Language

Language
Volume 94, Number 4, December 2018
Linguistic Society of America
Article

Viewed | Download | SaveSave

View Citation

Additional Information

The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change

Supplemental Material

Patrice Speeter Beddor, *Andries W. Coetzee*, *Will Styler*, *Kevin B. McGowan*, and *Julie E. Boland*

Related Article: https://muse-1jhu-1edu-100844fxq01a0.emedia1.bsb-muenchen.de/article/711824

[**Download HTML file**] The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change: Online supplementary materials

This HTML file contains the commented source code required to reproduce the analysis presented in the article in full, starting from raw output from the eyetracking, airflow, and duration measurement. This is an R-Markdown file and is best opened in R-Studio.

Patrice Speeter Beddor
University of Michigan

Andries W. Coetzee
University of Michigan and North-West University

Will Styler
University of California, San Diego

Kevin B. McGowan
University of Kentucky

Julie E. Boland
University of Michigan

Copyright © 2018 Patrice Speeter Beddor, Andries W. Coetzee, Will Styler, Kevin B. McGowan and Julie E. Boland

Additional Information

ISSN	1535-0665
Print ISSN	0097-8507
Launched on MUSE	2018-12-19
Open Access	Yes