1 Introduction

Coarticulation is phonetic variation due to the overlap of adjacent articulatory gestures. For example, to produce the word “bean,” the nasal gesture needed for the final /n/ begins early in the preceding vowel. We focus on anticipatory nasal coarticulation, since this type of coarticulatory variation is common across languages that have nasal consonants (approximately 98% of the world’s languages have a nasal consonant in their inventory, according to the World Atlas of Language Structures, Maddieson, 2013) and in every language with nasal consonants, some amount of coarticulatory vowel nasalization has been observed in production (Hajek, 2013). However, the theoretical backdrop of the current work, as well as the implications of the findings, are relevant for thinking about coarticulation more broadly. Coarticulation is a natural property of speech, and, in fact, necessary for fluent production. From an articulatory perspective, coarticulation is gestural overlap of phoneme units during speech production. Thus, coarticulation results in contextual variation in the realization of phoneme units. Coarticulatory variation is natural and ubiquitous, since it is an inevitable consequence of the influence that articulating one sound has on another (speech sounds are rarely produced in isolation!). However, there is ongoing debate and discovery about how coarticulation is represented in the mental lexicon of speakers. Furthermore, from a diachronic perspective, the pathway of CVN > CVN > C is a common historical change (Chen & Wang, 1975; Hajek, 1997). The goals of the present work are to (1) cover the key empirical and theoretical concepts relating to coarticulation, (2) demonstrate a laboratory phonological approach toward exploring and understanding patterns of coarticulation by presenting original and novel production and perception data, and (3) discuss the implications of investigating patterns of coarticulatory variation for the field of phonology. More broadly, we seek to examine the role of coarticulatory variation in synchronic and diachronic phonological variation.

1.1 Phonological Explanations for Coarticulatory Variation

One characterization of language is that it is a system that enables discrete elements to be combined in an infinite number of ways, with each combination conveying (or having the potential to convey) a different meaning. This applies to the phonological component of the grammatical system, as with others. For instance, “bed” and “ben” are lexical items composed of the same phonological elements except for the final one. Yet, these words also vary in that additional acoustic-phonetic cues to the final consonant are temporally distributed as coarticulation on preceding sounds. Because coarticulation is a type of context-dependent variation, whereby sounds take on predictable characteristics of other neighboring sounds, it might be
viewed as not relevant to encoding information critical to contrast between lexical items. For instance, some theoretical approaches aim to minimize phonetic-articulatory information that is stored in the underlying forms of words (Jackobson et al., 1952; Chomsky & Halle, 1968). In approaches such as these, the “idiosyncratic” information (i.e., minimal amount of symbolic-featural information that makes the word distinctive) and “systematic” information (i.e., contextually predictable details) hold different statuses in the grammar (Kenstowicz & Kisseberth, 2014). For a word like “ben,” the idiosyncratic information consists of the minimum amount of phonological detail that allows “ben” to contrast with the other legal word of English. Since there is no pair of words [bɛn] and [bɛn] that contrasts in nasalization on the vowel alone in English, these models assume that coarticulatory information is not stored: since coarticulation is predictable, it is classified as a type of systematic information that can be implemented post-grammatically.

Thus, since coarticulatory nasality is noncontrastive (since it does not encode information critical to discriminating between words like “bed” and “ben”), such a perspective might assume that its implementation in production is less important and less careful (perhaps weaker or more variable, etc.) than phonetic details directly located on the contrastive sound. Similarly, some have assumed that the perception of coarticulatory patterns is also not critical, or even that its presence could lead to perceptual confusion if it is not factored out or ignored by the listener; the acoustic consequences of the nasality might just make the vowel sound “bad” (i.e., less like the canonical oral version of the vowel) and thus harder to interpret (Lahiri & Marslen-Wilson, 1991).

More recent work, however, has shown that many noncontrastive properties of speech sounds are stored as well. In particular, research taking an experimental approach to phonological variation have explored how some properties related to the temporal implementation of phonemes are learned. An example of such an approach is Cho and Ladefoged (1999) who conducted a cross-linguistic survey of the phonetic implementation of voice onset time (VOT) in plosives across different place of articulation contrasts. They observed that differences in the temporal realization of VOT across languages are arbitrary. Since these temporal patterns were not consistent across languages and could not be tied to a strictly biomechanical explanation, Cho and Ladefoged concluded that such details had to be specified in speakers’ grammar as learned aspects of the phonetic representation of sounds. On the other hand, other aspects of the temporal realization of VOT, specifically, variation across place of articulation, were found to be consistent across languages and could be linked to an articulatory explanation. Their demonstration led them to argue for two levels of phonological encoding: a level containing segment contrasts (e.g., minimally specified phoneme units) and
a level containing language-specific phonetic values that are assigned to articulatory gestures (see Cho & Ladefoged, 1999, figure 10). In addition, automatic and universal articulatory implementation processes also apply during speech production, determined by physiological and aerodynamic processes. Thus, this approach demonstrates that precise experimental analysis can inform which phonetic differences are “phonological” (i.e., broadly meaning that they reflect learned and controlled variation), compared to those that are “mechanical,” thus not represented in the phonological grammar. Similar approaches to that taken in Cho and Ladefoged (1999) have been effective in outlining techniques for proposals about which coarticulatory features are physiological in origin and which are deliberate and targeted versus biomechanical aspects of the speech signal. Similar to the Cho and Ladefoged (1999) approach, there is much work demonstrating that patterns of coarticulation are language-specific (Beddor et al., 2002; Cohn, 1990; Keating & Cohn, 1988; Manuel, 1990; Schourup, 1973), indicating that speakers learn aspects of coarticulatory detail in their grammars.

What makes languages differ in coarticulatory features? Many researchers have searched for functional explanations for language-specific coarticulatory patterns. For instance, a prominent hypothesis for cross-linguistic differences in the production of coarticulation is that the system of contrast in a language constrains coarticulation (Manuel, 1990, 1999; Manuel & Krakow, 1984). The proposal is that when there is a lexical contrast to maintain for a particular feature (e.g., vowel nasalization), there will be less coarticulation (i.e., constrained coarticulatory vowel nasalization). For example, in French, vowel nasality is contrastive and encodes differences in word meanings (e.g., paix [pr] vs. pain [pɛ]). But noncontrastive coarticulatory nasality occurs in the context of nasal consonants as well (e.g., peigne [pɛɲ]) (Cohn, 1990; Delvaux et al., 2008). So oral vowels are nasalized when they occur adjacent to nasal consonants; this means that vowel nasality is both contrastive and noncontrastive in French. Thus, a listener who encounters a nasalized vowel in French has to discern whether the nasality is the result of a contrastively (i.e., phonologically) nasal vowel or an adjacent nasal consonant. In French, then, the fact that vowel nasality is contrastive in the language might limit the realization of anticipatory coarticulation, or might lead to perceptual difficulty or confusion to the extent that it does occur. Indeed, nasal coarticulatory patterns are observed to be less in degree and extent in CVN words in French, relative to that observed in other languages (Cohn, 1990; Delvaux et al., 2008; Montagu & Amelot, 2005). Hence, the differences in the patterns of contrastive and coarticulatory nasality can be viewed as representing perceptually motivated constraints on speakers in French.
However, studies of coarticulatory vowel nasalization in other languages with a nasal vowel contrast have found evidence for extensive coarticulatory nasalization, contrasting with the French case. Yet, they observe that the phonologically nasal vowel versus coarticulatorily nasalized vowel contrast can be maintained in other ways. For instance, Scarborough et al. (2015) examined both coarticulatory and contrastive vowel nasalization patterns in Lakota (a Siouan language of North America). They found that the location and magnitude of the nasalization peak within vowels varied across words in ways that are difficult to account for in biomechanical terms. Specifically, low phonologically nasal vowels in carryover contexts (i.e., with a preceding nasal consonant) exhibited peak nasalization at a late point in the vowel; meanwhile, in anticipatory contexts, the peak was at an early point of the vowel. Thus, phonologically nasal vowels which are in nasal consonant contexts and thus additionally subject to contextual nasality are produced with peak nasalization in the vowel portion furthest away from the coarticulatory source. Furthermore, peak nasalization in these contexts was of a larger magnitude than other phonologically nasal vowels; in fact, it is reported that some Lakota speakers describe the NV~ÑV as one of a “weakly” nasal versus “strongly” nasal vowel (Scarborough et al., 2015, p. 302). One explanation provided was that these variations are deliberate and derive from a system-wide need to keep the contrast between nasal and oral vowels distinct, especially in contexts where coarticulatory nasalization might neutralize that contrast. That native speakers are explicitly aware of the differences in magnitude provides support that this phonetic distinction is perceptible. However, an important point to note from the Scarborough et al. (2015) study was that they report individual variation in this observation: some speakers actually did exhibit neutralization of the NV~ÑV contrast. Thus, there appears to be variation within languages in the extent to which this type of contrast maintenance in nasal consonant contexts applies.

There are many other empirical studies from other languages that provide a nuanced picture of how coarticulatory patterns are constrained by the system of lexical contrast. In fact, there is some work exploring variation in coarticulatory patterns across languages, above and beyond what can be explained by systems of contrast. For example, Huffman (1988) compared degree and timing of nasal coarticulation in Akan (Niger-Congo language, Ghana) and Agwagwune (also Niger-Congo, Nigeria), two languages that differ in the phonological status of vowel nasalization. She found no differences in degree or extent of nasal coarticulation, though there was a difference in where peak nasalization occurred. Furthermore, Farnetani (1990) found nasal coarticulation in Italian to be very restricted, despite the fact that there is no oral-nasal
vowel contrast in that language. Aerodynamic analyses of nasal coarticulation in Bininj Kunwok (Gunwinngu language, Australia) also has shown restricted anticipatory nasal coarticulation, again despite the lack of a phonemic nasal vowel contrast in the languages (Stoakes et al., 2020). Thus, the cross-linguistic evidence suggests that any relation between phonological contrast and coarticulation may not be direct, especially with respect to nasal coarticulation.

A related question is, why is there systematic variation in coarticulation within a language. Some researchers have explored what factors might be relevant in determining within-language variation in patterns of nasal coarticulation. Similar to the Manuel (1990) hypothesis, many researchers have explored functional explanations for within-language coarticulatory variation. Lindblom’s body of work, for instance, takes the stance that phonetic variation is adaptive and the relationship between phonological organization and speech output is optimizing. For example, hypo- and hyperarticulation theory (Lindblom, 1990) proposes that articulatory variation reflects a dynamic, real-time trade-off between speaker-oriented and listener-oriented real-time communicative forces. The model proposes that phonetic variation reflects a speaker’s dynamic adaptation to communicative pressures. When the listening situation is ideal, the speaker can conserve articulatory effort and produce less effortful hypospeech. Yet, when there is some impediment to communication, the speaker might adapt their speech to be more intelligible to the listener by hyperarticulating. In effect, the H&H model predicts that the dynamic interplay between speaker- and listener-oriented forces in a conversational interaction will influence phonetic variation. Many of H&H theory’s predictions have been examined in empirical work on clear speech. For example, speech produced with the explicit goal to be clear exhibits longer segment durations and more peripheral vowel spaces, relative to less clear speech styles (e.g., Bradlow et al., 2003; Moon & Lindblom, 1994; Smiljanić & Bradlow, 2009). Furthermore, as predicted, speech with these phonetic adjustments has been shown to be more intelligible than conversational, or casual, speech (Bradlow & Bent, 2002; Picheny et al., 1985). Lindblom (1990) makes specific predictions about how coarticulatory patterns will manifest in response to trade-offs between the needs of the speaker and the listener. For example, his view is that the default “low cost” form of behavior by the motor system is increased coarticulation since it facilitates low-effort production of adjacent distinct articulations, while listener-oriented hyperspeech might contain reduced coarticulation in order to enhance the unique phonological properties of distinct segments. Yet, Lindblom (1990) does acknowledge that coarticulation might serve perception since it “does
provide valuable cues” for the listener (p. 425). Indeed, contrasting with Lindblom, Beddor (2009) posits that since coarticulation is perceptually informative in that it structures the speech signal to provide early and redundant information about the underlying features, it is something that speakers use in a highly controlled manner to provide cues to listeners. The notion that coarticulatory patterns are under speaker control and used to enhance coarticulation in strategic ways to help listeners comprehend and parse the speech signal has support from other work. Scarborough and colleagues (e.g., Scarborough, 2013; Scarborough & Zellou 2013; Zellou & Scarborough, 2015) explore how speakers’ patterns of coarticulatory vowel nasalization might be listener-directed in some cases. For example, Scarborough and colleagues have found that speakers systematically vary extent of nasal coarticulation across situations with more or less output-oriented pressures. For instance, in situations where there is an ostensible pressure to make speech more intelligible (e.g., when producing words that are more confusable due to a large number of phonological neighbors or in communicative tasks with an interlocutor where there is an authentic pressure to be more intelligible), speakers produce a greater extent of nasal coarticulation than in situations with less intelligibility oriented pressures (Scarborough, 2013). Degree of nasal coarticulation on vowels in French may vary systematically depending on neighborhood density, too (Scarborough, 2004). In other words, despite the fact that nasality is completely predictable and incapable of expressing contrastive information in English, and that nasality is contrastive as well as noncontrastive in French, nasal coarticulation is enhanced by speakers of both languages in contexts where CVN words are more confusable, potentially playing a role in perception. Parallel neighborhood-conditioned patterns have been shown for nonwords spoken by English speakers (Scarborough, 2012) as well as for other types of coarticulation, namely vowel-to-vowel coarticulation (Scarborough, 2004) and in tandem with other phonetic variables, particularly hyperarticulation (Munson & Solomon, 2004; Wright, 2004) and voice onset time (Baese-Berk & Goldrick, 2009). Thus, conditions where speakers have been shown to enhance primary phonetic cues to phonological contrasts, they also enhance anticipatory coarticulatory cues to upcoming segments.

1.2 The Role of Coarticulation in Diachronic Change

The fact that different languages exhibit distinct vowel nasality patterns has been the motivation for the development of theoretical models accounting for the role of coarticulatory vowel nasalization in historical phonological change. It has long been noted that synchronic phonetic variation is a precursor to the
emergence of innovative sound patterns over time in a speech community (Weinreich et al., 1968). Since coarticulation results in natural and systematic types of synchronic variation in the realization of phonemes, its role in diachronic sound changes has been extensively explored. For instance, coarticulation results in an ambiguous relationship between the speech signal and its phonological structure: [ṼN] could be interpreted as signaling either /VN/ or /ṼN/, or some other possibility, as the intended underlying form. Thus, from the listener’s perspective, coarticulation creates a one-to-many (or, many-to-many) mapping between the acoustic signal and the underlying structure of the utterance. The process of a listener “backwards-engineering” the speaker’s gestural behavior from the speech signal has been hypothesized to be a mechanism for diachronic sound change (e.g., Ohala, 1993). For instance, Ohala’s (1993, inter alia) listener-based model of sound change origins is one prominent proposal for how speech processing mechanisms lead to the phonologization of coarticulation in this way. Ohala argues that listeners generally normalize for coarticulation by ascribing the acoustic effects of gestural overlap to their sources in the speech signal. For nasal coarticulation, this would mean that listeners attribute the vowel nasality in the context of a nasal consonant to its source, and thus analyze that the vowel as inherently nonnasal (e.g., [ṼN] is “reverse-engineered” to /VN/). Evidence for this process comes from work examining how perception of nasal coarticulation depends on context: English listeners hear a nasalized vowel in isolation as nasalized, yet they hear a nasalized vowel in the context of a nasal consonant [NṼN] (Kawasaki, 1986).

However, another possible interpretation of [ṼN] is as /ṼN/. In this scenario, a listener perceives the nasality in the speech signal as an intended property of the vowel and interprets it as reflective of the speaker’s underlying representation for that utterance. According to Ohala, in this case, if the speaker’s original representation for that utterance was an inherently nonnasal vowel, the listener has “hypocorrected” or “failed” to normalize for the effect of context. The hypocorrection scenario is an event where the underlying representation for a given utterance is different for the speaker and the listener. Such an event has been described as a mini sound change, and accumulation of such situations is hypothesized to be the origin of long-term changes in production targets for speech sounds. However, listeners’ sensitivity to coarticulatory variation on sounds in context have been demonstrated in many studies of “partial” perceptual compensation (e.g., Beddor & Krakow, 1999; Fowler, 2005). For instance, Beddor and Krakow (1999) use vowel discrimination tasks where listeners hear two versions of an NVN lexical item, differing only in the presence of nasality in its vowel: one token contained a nasal vowel and the other contained an oral vowel. Participants’ responses in discriminating between vowels that contain
nasalized vowels in context are above chance, indicating that some coarticulatory effects routinely remain perceptible by listeners (Beddor & Krakow, 1999; Krakow & Beddor, 1991). This is evidence of “partial” compensation for coarticulation and it is argued that sound change might arise via phonologization of this residual coarticulation.

Partial compensation for coarticulation as a potential sound change mechanism has been recruited to explain many common sound changes observed in connection with vowel nasalization. For example, nasalization of vowels results in an acoustic-perceptual lowering of F1 (Krakow et al., 1988). A listener might misattribute the lowered acoustic center of gravity due to the spectral properties of nasality not to the source nasal lowering gesture, but instead to a raised tongue position (which would also lower F1). Thus, the acoustic ambiguity of nasalized vowels presents opportunities for sound change. Indeed, low vowels with nasality are perceived by listeners as higher than oral low vowels (Krakow et al., 1988; Wright, 1986). Furthermore, speakers systematically produce distinct oral articulations for nasalized vowels (both contrastively and coarticulatorily nasalized) relative to nonnasalized counterparts in many languages (Carignan, 2014, 2017; Carignan et al., 2011). This has been argued to reflect the phonologization of distinct oral and nasal vowel qualities that originated in the misattribution of nasalization effects as tongue height variations (Carignan et al., 2011). The effects of nasalization on perceived vowel height can explain historical vowel changes that occur in nasal consonant contexts, such as the raising of /æ/ in CVN words (Boberg & Strassel, 2000).

Another mechanism for sound change in the classic Ohala model is hypercorrection. Another speaker, whose representation perhaps reflects previous phonologization of the acoustic effects of nasalization (i.e., someone who has phonologized CVN as /CṼN/), produces an acoustic signal that more closely resembles their intended utterance: [CṼN]. Yet, the listener might make the same assumption for correction, that the nasality is the result of mechanical coarticulatory processes and not intended by the speaker. Thus, the listener might attribute these features to the adjacent nasal coda, “over-correcting.” In this case, the speaker’s intended representation of the utterance (with a deliberately nasalized vowel [CṼN]) and the listener’s interpreted representation of the utterance (with an oral vowel, since they have subtracted the nasality in the signal as due to mechanical processes [CVN]) do not match. There is some empirical evidence that such a process occurs (although it is less well examined than hypocorrection). For example, Harrington et al. (2013) found that listeners compensated less for coarticulation in unstressed vowels, which they interpret as an instance of hypercorrection. In this case, the speaker’s
intended representation of the utterance and the listener’s interpreted representation of the utterance also do not match, like hypocorrection, yet the listener has normalized for coarticulation to a greater extent than the speaker intended. One of the phonological implications of over-normalization is that it could lead to dissimilatory sound changes. It is interesting to note that changes involving hypercorrection are much less often explored in the sound change literature (a question that is open to be explored in future work).

Ohala’s model frames the mechanisms for sound change as resulting from “misperception.” This assumption has been questioned by recent researchers examining sound change and revised models of phonologization of coarticulation have been proposed that address these points. In particular, Beddor’s (2009) model of a coarticulatory pathway to sound change relies on many of the same assumptions as the previous model, with some important modifications. Beddor (2009) views coarticulation as perceptually useful information, emphasizing the importance of anticipatory nasal coarticulation in providing early cues to the listener to help comprehend the speech signal. For example, anticipatory nasalization can be used by the listener, just by hearing [bɪ], to identify that the word is “bean” not “head” (Beddor, 2015; Beddor et al., 2013; see also Zellou & Dahan, 2019). Thus, Beddor’s model assumes listeners routinely use coarticulatory detail as part of the process of efficiently comprehending the speaker’s message.

Furthermore, Beddor (2009) integrates both the speaker’s and the listener’s roles in explaining how synchronic variation might lead to sound change. For example, coarticulatory variation is a product of speaker-controlled deliberate trade-offs in production. This was examined in a production study of American English words with ~NC sequences (such as “bent” and “bend”) where the nasal consonant typically shortens (or deletes) before a final voiceless oral coda. Yet, she observed that while there was variation in the duration of the nasal segment, the total temporal extent of nasalization remains identical across words like “bent” and “bend” indicating the speakers are actively maintaining the duration of nasalization present across these words. Even further, Beddor (2009) argues that the speaker deliberately maintains the velum-lowering gesture even when a shorter gesture might be less effortful. Thus, Beddor (2009) frames this as both grammatical and perceptually oriented: speakers are strategic by realigning the onset of nasalization so that the same amount of that cue is present in the acoustic signal even if it is temporally or spatially realigned. Beddor (2009) links both speaker-controlled and listener-oriented nasal coarticulatory patterns to sound change. Under such a speaker- and listener-driven sound change view, patterns of nasal coarticulation are a learned part of the phonetic structure of the language, used by speakers in strategic ways and the object of close attention by
listeners. Indeed, listeners are highly sensitive to the acoustic effects of coarticulatory nasalization in discriminating lexical contrasts in English. For instance, Ali et al. (1971) found that when a final nasal or nonnasal consonant was spliced out from a monosyllabic word, listeners could reliably predict the missing segment.

Thus, Beddor (2009) proposes that listeners’ veridical attention (i.e., sensitivity to coarticulatory details) to the speech signal gives rise to sound changes. Specifically, she hypothesizes that perceptual equivalence between the multiple possible interpretations for a given utterance, both within and across individuals, is the mechanism for sound change. This was demonstrated when she looked at the perception of VNC sequences with manipulated temporal extent of nasal coarticulation. She found that listeners, when rating words manipulated to vary in the relative duration of the nasalized portion of the vowel and the nasal consonant, attend to the total sum of nasality duration across a syllable, rather than the extent of nasality localized on one particular segment. Hence, listeners do accurately hear information about nasal coarticulation present in the acoustic signal; but due to the ambiguous relationship between the speech signal and the multiple possible underlying gestural mappings, listeners have multiple options to decide on the intended utterance. Thus, variations in the alignment of the nasalization gesture can lead to perceptually equivalent realization—the listener’s choice in selecting which representation reflects the acoustic signal might lead to variants that are then emphasized by listener-turned-speakers. In this case, veridical perception of coarticulatory details is not misperception, rather since there are multiple possible and appropriate interpretations for a given VN sequence, listeners’ choices are not “wrong” (see also Lindblom et al., 1995, who reframe hypocorrection as an adaptive mechanism that is sensitive to semantic and situational factors), but reflect differences in the relative weighting across listeners of coarticulation as a cue to word perception.

Another model that revisits and extends some of the hypocorrection model ideas is Blevins’ (2004) model of sound change typology. Blevins also views synchronic phonological variation as partially reflective of historical sound change. In particular, Blevins looks for explanation in typological patterns: the fact that some phonological alternations are common and others are rare can be explained in terms of how likely or not a respective sound change is to occur. Blevins’ typology emphasizes understanding the aerodynamic, acoustic articulatory, and perceptual motivations for sound change. Blevins (2004) argues that sound change is a result of the nature of linguistic representation. For instance, Blevins assumptions about lexical storage are based on exemplar models (Pierrehumbert, 2002, 2016). Blevins (2015) argues that the proposal that lexical representations include memory traces of words with experienced