Prospectus: Modeling the interface between phonetic biases, synchronic category learning, and language change for explaining asymmetries in the typology of vowel nasalization

Motivation

We often find parallels between phonetic experimental results and cross-linguistically frequent contrasts, processes, and diachronic sound changes. For example, nasal contrasts are more frequent for low vowels than high vowels, and in production data, low vowels are produced with a lower velum, associated with greater nasality. These parallels often lead us to hypothesize that the differences found experimentally *cause* differences in cross-linguistic frequency. However, we do not yet understand, or even have detailed proposals for, the linking mechanisms that would allow synchronic phonetic differences observed in speakers at one point in time to affect the development of diachronic sound changes, and in turn, the later language systems. Specifically, we do not yet understand how differences in perception and production interface with how people learn the categories and processes of their language, and of particular relevance to sound change, how speakers learn a different system than the one that generated their learning data (the previous generation's). Furthermore, how speakers learn their languages' categories and phonological processes in general, aside from any phonetic biases, is itself an open question and area of ongoing research.

In this dissertation, I will develop and implement hypothesized models of phonetic category learning, focusing on how learning interfaces with production and perception differences. Expanding to agent-based modeling and using experimental phonetic data, I will jointly evaluate learning models and hypothesized phonetic biases with typological frequencies. By making concrete, implemented hypotheses about linking mechanisms, I will test claims about the type and size of phonetic differences needed to affect typological frequencies. I will focus on a case study where parallel typological differences and phonetic asymmetries have been observed: vowel height categories in nasal vowel inventories. However, I intend the hypothesized category learning and sound change mechanisms to be general and therefore applicable (and testable) in other cases of typological asymmetry in contrast frequency, e.g. stop voicing contrasts in different contexts.

In explaining the cross-linguistic frequency of sound processes with phonetic parallels, researchers have long debated why phonetically-related processes and contrasts tend to be typologically frequent. Some have argued that, synchronically, language learners are more likely to learn phonetically-related patterns because of cognitive learning biases in favor of them (e.g. Wilson 2006, Flemming 2003). Others instead focus on diachrony, arguing that phonological learning is abstracted away from phonetics with no such biases, and the only reason that phonetically-related processes are common is that sound change is sensitive to production and perception (Blevins 2004). These two possible influences on phonological typology have been described as analytic biases (differences in learnability, from cognitive systems) and channel biases (differences in the likelihood of sound change, from production/perception); work on differentiating them includes artificial language learning experiments (Moreton & Pater 2012, Wilson 2006, Glewwe 2022), estimations of the size of phonetic differences (Moreton 2008) and the probability of different sound changes (Beguš 2021), and evaluating the learning biases of models with different featural or constraint representations (O'Hara 2021). However, neither perspective has yet proposed a complete account of how mechanisms at the individual and population levels (such as in speech perception, category learning, and language transmission) would allow synchronic phonetic factors (learning biases or otherwise) to influence diachronic sound change, and in turn which contrasts and processes are likely to be established in languages. Extensive work in phonetic and phonological learning (e.g. Maye 2008, McMurray et al 2009, Feldman et al 2013) and recent work in the computational modeling of sound change (Soskuthy 2013; Kirby 2014; Todd 2019; Harrington and Schiel 2017; Gubian et al. 2023) provide a basis for exploring more concrete accounts. By making explicit, implemented proposals for the mechanisms linking experimentally-supported phonetic differences to typological differences, this dissertation will further enable each perspective to develop testable predictions.

There have been verbal proposals about listeners' misattribution of speaker intention (e.g. Ohala 1994; Beddor 2009; Barnes 2002), but none have been implemented in detail computationally. There are a number of implemented exemplar-based proposals for how categories shift, merge, or maintain contrast (Wedel 2004; Blevins & Wedel 2009; Soskuthy 2013; Kirby 2014; Todd 2019), but not split into separate phones, which is necessary for understanding the development of phoneme inventories (Labov 1994). One model that does make a proposal for category splitting (Gubian et al. 2023) has not yet been applied to phonetic naturalness biases and might make unsupported predictions about the development and transmission of category splits, discussed in Chapter 3.

In this dissertation, I will use the case study of nasal vowel inventories to make concrete proposals about the interaction between category learning, phonetic asymmetries, and sound change, and test some of their predictions. The asymmetry in nasal vowel inventories provides an informative test case because multiple phonetic parallels have been found in production and perception, but it is not clear which (if any) predict the typological asymmetry; considering them jointly with learning hypotheses may help distinguish them. In addition, the development of vowel nasalization is an instance of one sound category (an oral vowel) splitting into two (oral and nasal variants). Although vowel nasalization is a difficult phenomenon to study articulatorily (involving the velum and nasal cavity), acoustically (lacking a well-defined measure independent of the oral cavity resonances), and perceptually (designing tasks to estimate listeners' perception of vowel nasality independent of consonant nasality), there is a base of experimental work on vowel nasality designed with sound change and language-specific differences in mind (e.g. Beddor 2009, Kunay et al. 2022, Carignan et al. 2021). Acquiring a language with or without vowel nasalization is also an interesting learning problem by itself: inferring whether a feature (nasality) perceived during a sound (vowel) is due to a target for that sound, or an accidental effect of the neighboring context. Although the learning model I've been working on so far doesn't frame the task in this way, it's a direction for further work on developing a sound category learning model. The typology of nasal vowel contrasts also bears on the question of the mechanisms behind the typology of vowel contrasts in general, which is a large area of ongoing work (e.g. Schwartz et al. 1997, de Boer 2001,Trudgill 2009, Flemming 2017, Vaux & Samuels 2015).

Background

Section 1. Nasal vowel typology

Cross-linguistically, nasal contrasts are more common for low vowels than high vowels; out of an UPSID sample of 102 languages with nasal vowels, 20 are missing at least one high nasal vowel compared to only 6 that are missing at least one low nasal vowel (Kingston 2007, cf. Chen 1975, Ruhlen 1975). For example, Amuzgo contrasts [a] and [ã] but has no nasal counterpart to its oral [i] (Longacre 1966).

Multiple interactions between height and nasality have been found in experimental work, and these interactions have each been discussed as a potential cause for the higher frequency of low nasal vowel contrasts. However, the effect each of these interactions has on synchronic learning, diachronic sound change, and language development remains unclear. In this dissertation, I will explore concrete learning and sound change mechanisms to generate testable predictions for two main accounts of the higher frequency of low nasal vowel contrasts: (i) more velar opening for low vowels and (ii) greater perceptibility and/or production of nasality for vowels of longer duration, which is associated with vowel lowness (Henderson 1984; Whalen & Beddor 1989; Hajek 1997; Kunay et al. 2022).

Section 2. Experimental phonetic parallels to typology

Section 2.1. Velum lowering in low vowels

For vowels of different heights and contexts, Henderson (1984) measured velar height and port opening, and measured muscle activity with electromyographic potentials (levator palatini, palatoglossus, superior pharyngeal constrictor). Measurements were taken from the productions of speakers of a language without phonemic vowel nasalization (English) and a language with a nasal-oral contrast for every vowel (Hindi). Across both languages and all contexts (neighboring nasal consonant, neighboring oral consonant, phonemically nasal vowel, phonemically oral vowel), low vowels had significantly more velar opening than high vowels. Henderson (1984) suggests this greater nasality of low vowels in production as a possible cause of the greater cross-linguistic frequency of nasal contrasts for low vowels, perhaps because oral low vowels are more similar to nasal vowels. This result has since been replicated, e.g. by Kunay et al. (2022).

However, it is unclear how the nasality of low vowels would increase the likelihood of a split into the oral and nasal categories involved in a contrast. The potential relation between low vowels' velar production and nasal contrasts has also been referenced by Whalen & Beddor (1989), Blevins & Garrett (1993), and Barnes (2002). Hajek & Maeda (2000) argue against this production difference as the cause of the greater frequency of low vowel nasal contrasts, not based on the linking hypotheses for how greater nasality would spur contrast, but based on the strength of empirical evidence for a production interaction between vowel height and nasality: some studies have failed to replicate the relationship (Clumeck 1976) and greater velar opening in production does not necessarily result in a greater perception of nasalization for low vowels (Maeda 1993).

Section 2.2. Height, Length, and Nasality Interaction

Whalen & Beddor (1989) synthesize vowels with a constant degree of velar port opening and varying duration, and find that listeners' nasality ratings interact with vowel duration, with more nasal responses for longer vowels. This relationship holds for vowel stimuli synthesized with different degrees of velar port opening. However, longer duration did not increase nasal responses for unambiguous natural tokens or for unambiguously oral tokens created by copying pitch periods from natural oral vowels. For stimuli spliced from natural productions of vowels, duration only interacted with nasality for tokens with some pitch periods spliced from a nasal vowel. In summary, increased duration is correlated with increased nasal responses specifically in ambiguous cases where there is some degree of vowel nasality. However, Whalen & Beddor (1989) did not examine the effect of neighboring consonant context.

Cross-linguistically, low vowels tend to be longer than higher ones (Peterson & Lehiste 1960; Lindblom & Rapp 1973). These interactions between height and length, as well as length and nasality, suggests the cross-linguistic preference for nasalizing low vowels could result from low vowels being perceived as more nasal due to their duration. Whalen & Beddor (1989) argue that, together, the sizes of (1) the difference in length between low & high vowels, and (2) the effect of duration on nasal perception, are both too small to be the cause of the cross-linguistic preference to nasalize low vowels. It is unclear how large the effect size would need to be to affect the typology (as previously noted by Hajek 1997), especially when we do not yet understand what learning and transmission mechanisms would be affected by the nasality perception bias; Smith et al. (2017) shows that agent-based language change models can predict large typological differences from small biases, and vice versa. This dissertation will take a first step toward clarifying the relation between phonetic nasality biases and the development of vowel nasalization.

Section 3. Models of category learning and sound change

Exemplar models with Mixture of Gaussians learning are often assumed and tested in the category learning (Pierrehumbert 2001, Maye 2008, Feldman et al. 2013) and sound change literatures (e.g. Todd 2019, Kirby 2014, Soskuthy 2013, Morley 2013, Gubian et al. 2023). A Mixture of Gaussians (MOG) learner searches for the set of categories (each a multivariate Gaussian distribution) that maximizes the likelihood of its input data. Intuitively, the less overlap between vowel distributions, the more likely they'll be learned as separate categories (cf. Feldman 2013).

For modeling the development of vowel categories, this dissertation assumes a learning model that has to infer the number of categories in its input (so that there may be variation in the number of categories across learners and generations). Although traditional Mixture of Gaussians learners assume a fixed number of categories as a hyperparameter (e.g. Soskuthy 2013), two possible approaches used in the category literature allow them to learn a category set of any size: (1) running MOG learning multiple times, with a different number of categories, and using model comparison (e.g. with BIC) to select the best number of categories (used in Gubian et al. 2023) or (2) replacing the prior probability over each category with a Dirichelet process, which can generate a new category to assign a token to, with some probability depending on the token's similarity to the existing categories as well as the number and size of existing categories (used in Dillon et al 2011 and Feldman et al 2013). When a learner's input is noisily sampled from parents' categories, it may learn a different set of categories; for example, it may split a single category into two.

I am in the process of exploring the hyperparameter space of the Dirichlet process Mixture of Gaussians learner to find values that allow for consistent convergence and some variability in the number of categories learned given input sampled from the same parent distribution. I am trying different sizes of the concentration prior over the number of categories, the size of the learning input, and optimization details (e.g. the maximum allowed number of learning iterations).

Many predictions of the sound change model in this dissertation rely on the property of Mixture of Gaussians learning that the degree of overlap between two distributions in the learning input will affect the likelihood a learner will infer one category or two (Feldman et al. 2013). Although the properties of Dirichlet Process Gaussian Mixture Models have been explored for accuracy under different conditions (Steinley & Brusco 2011), I have not found this intuitive property of mixture models demonstrated analytically or with systematic simulations (although see Feldman et al. 2013 for a comparison of model performance on a less-overlapping dataset and more-overlapping dataset). So, as a sanity check of this property of the model, I am also using simulations to empirically test the intuition that the degree of overlap in the parent categories affects the likelihood of the child learner splitting or merging categories, across different hyperparameter values.

Section 4. Previous work on sound change pathways

Based on diachronic analyses across multiple languages (Ferguson 1963; Hajek 1997), the development of nasal vowel contrasts has been assumed to involve two processes: first, a language without nasal vowels develops allophonic vowel nasalization before nasal consonants (e.g. a → a and an → ãn). Additionally, the language undergoes a sound change with final nasal consonant deletion, so vowel nasalization becomes contrastive by itself (e.g. a → a and ãn → ã) (Ferguson 1963).

The time course of these two changes is an open question: based on synchronic phonetic data on the tradeoff between vowel nasality and nasal consonant duration in American English and German, Beddor (2009) and Carignan et al. (2021) hypothesize that both steps occur simultaneously. Hajek (1997) instead hypothesizes that the development of allophony and the loss of the nasal consonant occurs at two separate timepoints, based on existing contrasts and historical reconstructions for varieties of French and Italian. This hypothesized two-step sequence predicts that low vowels would be more likely to have a nasal contrast if they are more likely than high vowels to split into oral and nasal allophones (the initial a → a and an → ãn sound change).

Distinguishing the predictions of these two hypothesized time courses might be outside of the scope of this dissertation; for now, I assume the two-step time course in the implemented hypotheses. The reason for choosing this time course instead of the simultaneous one is to have a simpler starting point: I can focus on sound change hypotheses in terms of only vowel nasalization, rather than both vowel nasalization and nasal consonant deletion at the same time.

Section 5. Coarticulation vs allophony

In assuming the two-step time course, an objective of this dissertation is to design and test models of how learning and language transmission result in allophony, specifically vowel nasalization. These models would be tested jointly with the influence of biases that would result in allophony more often for low vowels than high vowels.

A resulting task is to operationalize the distinction between nasal vowel coarticulation, nasal vowel allophony, and nasal vowel contrast. However, the boundary between coarticulatory processes and allophony is not always clear: is the influence of a neighboring sound solely due to physical/timing constraints, or is it intentional and learned language-specifically by the speaker? How big is the influence of a neighboring sound? I assume (i) nasal vowel coarticulation is limited, not learned, and a consequence of physical/timing constraints on velum lowering (ii) allophony reflects learning of the difference between oral/nasal vowels and a vowel nasalization target for vowels preceding a nasal consonant, and (iii) a nasal contrast reflects a learned oral/nasal vowel distinction that carries lexical meaning by itself, without the presence of a nasal consonant. Work on this topic has potentially wider implications because several other phonological allophony and sound change processes have parallels in coarticulation, e.g. vowel harmony (Rysling and Kingston 2024) and velar stop palatalization (Wilson 2006, Morley 2013).

Solé (1995) finds distinct patterns of nasalization in American English and Spanish: the American English speakers produced long durations of vowel nasalization, with the nasalized percent of vowel duration constant across speech rates, while the Spanish speakers produced very short durations of vowel nasalization, with the nasalized percent of vowel duration variable across speech rates. Solé argues that this difference in duration and variability of vowel nasalization represents a difference in phonology: American English speakers have a nasal target for vowels in a VN context, resulting in the long and consistent nasalization, whereas Spanish speakers don’t, resulting in the short and variable nasalization in a VN context. Therefore, in the assumed operationalization of the difference between coarticulation and allophony, American English has nasal vowel allophony, whereas Spanish has just nasal coarticulation.

However, there is also individual variability within American English: Beddor (2009, 2018) finds differences in the extent to which individuals use vowel nasalization in production and perception, and Zellou (2022) finds three distinct patterns. Based on these experimental results, there might be two relevant dimensions for the influence of a nasal context: whether there is a learned, stable target particular to sounds in that context (allophony versus coarticulation) and how different that target is from other contexts (different realizations of allophony).

I hypothesize that language-specific and individual-specific differences in vowel nasalization come from differences in exemplar category representations: allophony is the result of two distinct exemplar clouds/categories for vowels in oral versus nasal contexts (as opposed to a single shared exemplar cloud with a shared mean and variance), and different realizations of allophony are the result of different distances between category distributions. Specifically, one possible formalization of coarticulation versus allophony is that allophony is defined by speakers learning two separate categories (e.g. a Gaussian distribution for the vowel in [bad] and a Gaussian distribution for the vowel in [ban]) whereas unintentional coarticulation is so small and variable that speakers represent phones in both contexts with a single Gaussian category (a single Gaussian shared by both [bad] and [ban]). By exploring and testing this model of category learning involved in nasal typology, this dissertation also bears on our understanding of allophony.

Chapter 1: Typological predictions of velum lowering and Dirichlet Process MOG learning

Overview

I demonstrate that, when assuming an often-assumed (e.g. Kirby 2010; Gubian et al. 2023) model of category learning and sound change, low vowels' production bias toward greater nasality (Henderson 1984) does not predict they are more likely to split into nasal/oral contrasts.

Section 1.1. Definition of the simple Dirichlet Process Mixture of Gaussians (MOG) model used in simulations

1i. Mixture of Gaussians Model setup

Simulated vowel tokens are defined on two dimensions: vowel height and vowel nasality. An individual's phonetic categories are defined as two-dimensional Gaussian distributions with means μ_k, covariance matrices Σ_k, and prior probabilities $\boldsymbol{\pi}_k$ (determined by how large the category is, i.e. how likely tokens are, on average, to belong to that category based on μ_k and Σ_k) An individual's learning input consists of a collection of vowel tokens they have heard, of unknown categories. Tokens in the learning input are assumed to be defined on perceived vowel height and nasality, and tokens sampled from an individual's categories are assumed to be defined on produced vowel height and nasality.

Categories are learned in a Bayesian version of Expectation-Maximization, where the number of categories is learned with the "stick-breaking" approximation of a Dirichelet process, with a hyperparameter for the maximum possible number of categories. This algorithm is implemented by SciKit-Learn's Bayesian Gaussian Mixture class. A core intuition of Mixture of Gaussians models is that the more overlap between two distributions, the more likely they are to appear unimodal, and the more likely they are to be learned as a single category (cf. Feldman et al. 2013); conversely, the less overlap between two distributions, the more likely they are to be learned as two separate categories. Even when a learner has a single category representation for oral-context and nasal-context vowels, nasal coarticulation is operationalized as a bias that makes vowels produced in a nasal context sampled from a slightly higher mean nasality.

1ii. As a model of sound change

An individual's learning input consists of their parent's productions: a sample from their parents' Gaussian category distributions, where the individual is unaware of which parent category each vowel came from. Because this sample is noisy, an individual's learning input is not identical to their parent's learning input, so it's possible they will infer a different set of categories.

More complex learners have been used as a model of category learning in sound change for tonogenesis (Kirby 2014), the effect of phonetic biases given a number of categories (Soskuthy 2013), allophony (Dillon et al. 2011), sound shifts/mergers (Gubian et al. 2023, Todd 2019), and in one case, a fricative split and merger (Stevens et al. 2019). However, all of these models' agents use Mixture of Gaussians for category learning, even if they have additional processes (e.g. an incentive for contrast maintenance and/or an extra layer of sub-category Gaussians for lexical items or contexts). By starting with a very basic implementation of Mixture of Gaussians learners, I can test predictions that are agnostic of these variations and likely common to them; for example, the effect of distributional overlap on category splits. One caveat is that, as far as I know, nobody has shown that all of these Mixture of Gaussians variations share this relationship between distributional overlap and category splits. It would be surprising if they did not, since they share the learning objective of finding the set of categories that maximizes the likelihood of the data, without having more categories than necessary. However, if I have space within my dissertation timeline, I would like to verify this intuition through running simulations for each of these more complex models (many of which have publically available code) where I vary the degree of overlap in the starting distributions and measure how many categories are predicted across simulations (because there is stochasticity in each simulation).

1iii. Sound change pathways and synchronic stages

Following the sound change pathways proposed by Hajek (1997), I have outlined the progression of synchronic stages and sound changes where a language could start with no nasal vowel contrasts, and then proceed to having nasal contrasts, either missing a high nasal vowel, missing a low nasal vowel, or having all nasal vowel contrasts. This schema can be extended to mid nasal vowels, which are excluded here for simplicity since the main focus is on the difference in typological frequency between high and low.

I show what individuals' category representations would look like at each stage. In order to transition from one stage to another, learners need to "mislearn" the productions from the previous stage as having come from the categories of the next stage, as shown in Figure 1. Following the operationalization of allophony defined in background section (e), individuals with nasal allophony for a given vowel have 2 Gaussian categories instead of one for that vowel.

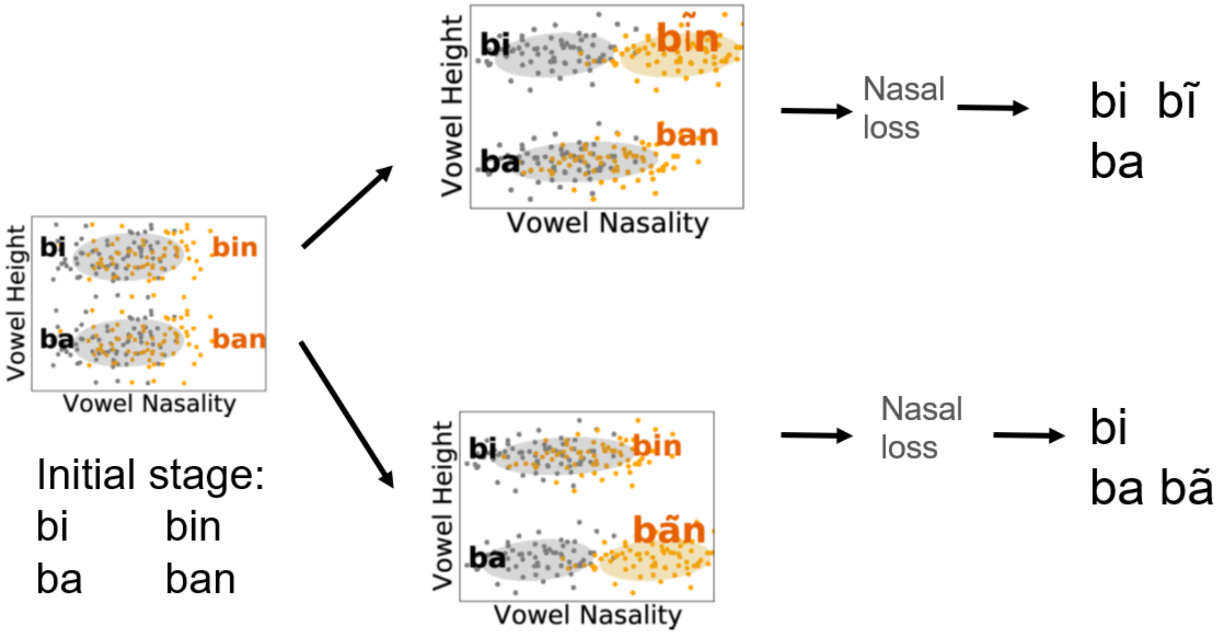

Figure 1. Sound change pathways and corresponding category representations for individuals at each stage, for a toy lexicon starting with no nasal vowel allophony or contrast.

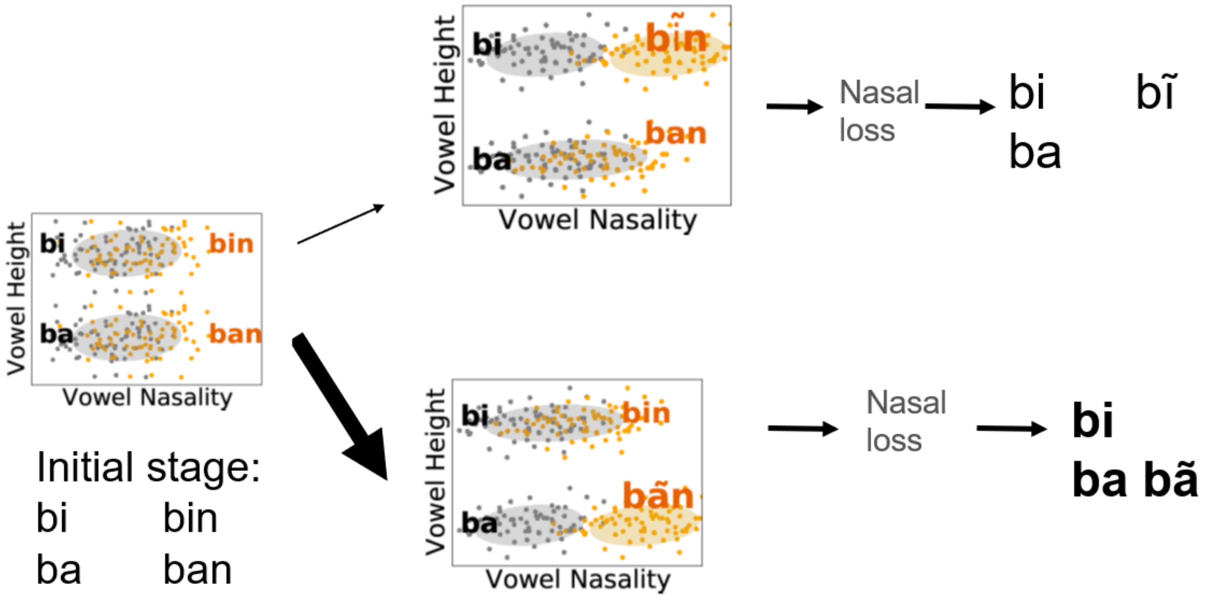

Figure 2. The pathway at the bottom is highlighted with increased arrow weight because a model would need it to be more likely in order to correctly predict the asymmetry in contrast frequency.

As a reminder, for simplicity, I am assuming Hajek (1997)'s hypothesis that nasal vowel contrasts start as nasal vowel allophony, then later result in contrast when an additional sound change results in nasal consonant loss. In order to predict the typology that nasal contrasts are more common for low vowels, the sound change pathway resulting in nasal vowel allophony must therefore be more common for low vowels (Figure 2). One empirical prediction these proposed sound change pathways make is that nasal allophony frequency, not just nasal contrast frequency, will be greater for low vowels. The Phoible database, alongside its inventories of (hypothesized) phonemes, includes a record of their allophones. Because this difference between allophone frequencies would therefore be straightforward to estimate, I will conduct this analysis to further evaluate the evidence for this model of vowel nasalization sound change.

Other sound changes are possible; for example, vowels may later merge in nasal contexts. For simplicity, they are not included here, but if there's time, I would like to include them in this hypothesis space of pathways to different nasal vowel inventories.

1iv. Implementing bias based on previous articulatory results

As the model has been defined so far, there is nothing that will make the nasal vowel allophony pathway more likely for low vowels (Figure 2).

Henderson (1984) hypothesizes that low vowel nasal contrasts are more frequent because of low vowels' greater velar lowering. To establish whether MOG learning is consistent with this hypothesis, I implement it and show whether it correctly predicts that nasal contrasts are more common for low vowels; that is, whether it makes the low vowel sound change more likely than the high vowel sound change (Figures 1 and 2). I implemented this production bias by shifting all simulated productions of low vowels to a higher value on the nasality dimension by a constant amount (as Henderson (1984) found low vowels are more nasal than high vowels across all contexts). The simulation was set up in this way:

I defined starting point categories for the first generation individual. As a first attempt at implementing the hypothesized low vowel velum lowering bias, low vowels are sampled from a higher mean nasality (+10). The units are arbitrary for the abstract height/nasality dimensions.The covariance matrices for low and high were defined identically, with a positive correlation between the height and nasality dimensions.

	Mean height	Mean nasality
High	250	110
Low	150	110+10=120

Table 1. Summary of mean height and nasality values for the high and low vowel categories in the input distribution.

To operationalize the effect of coarticulation, production samples for the [b_n] context were sampled from a Gaussian with a slightly higher (+10) mean from each category. I sampled 250 productions each from these production distributions (bid, bin, bad, ban). I then used the unlabeled production samples as training data to the MOG model for the next generation. If the low vowel nasal contrast pathway is more likely under this model, then across runs (since there's stochasticity) the learner should be more likely to learn separate nasal/nonnasal categories for low vowels than high ones. Two hyperparameter settings were tried for the maximum number of categories: 4 (the maximum number of categories if the learner recovers all of the possible sub-distributions that generated the data) and 10 (to be more realistic to the number of possible vowel categories a learner might consider)

Across runs, it did not seem low vowels were more likely to split; either both categories split, or neither; only rarely did one of the vowel heights split, and it appeared to occur equally often for high and low vowels. I am currently performing these simulations systematically, by collecting the number of times low categories versus high categories were split; these simulations will be run across multiple random seeds and hyperparameter values to evaluate how big the difference is related to the noise in the learning simulations.

This preliminary result is an unsurprising finding because the amount of distributional overlap between vowels in nasal and oral contexts (ba vs ban) will still be the same if all values are shifted by the same amount, no matter the size of the shift, and the separation or merger of categories in Mixture of Gaussians depends on the amount of overlap in the distributions that generated them.

An additional concern was that the definition of Mixture of Gaussians used in SciKitLearn (Bayesian, with an approximation of a Dirichelet prior for generating new categories) is sensitive to the number of samples (more samples result in generating more categories overall); personal correspondence with those working on the Gubian et al. (2023) model, which used a different infinite MOG implementation (running non-infinite-category MOG with different category numbers, then picking the number of categories based on BIC model selection), showed they also encountered this behavior. However, by systematically varying the number of learning iterations and number of samples, I discovered that this tendency toward a greater number of categories only holds when the number of learning iterations is low relative to the size of the learning input, resulting in the model failing to converge on some runs. In the rest of the simulations I am running, I ensure the number of iterations is high enough that the number of categories learned does not grow rapidly with the number of learning input tokens, and that the model reliably converges.

Remaining work, beyond establishing a more rigorous statistical test of this result as discussed above, is to establish the predictions of this model and bias over more generations of learners, not just 2 generations (parent and child), as population structure matters for the predicted frequency of sound changes (Smith et al. 2017; Kirby & Sonderegger 2015; cf. Soskuthy 2013). Chapter 3 will focus on this investigation of larger population structures.

Section 1.2. Empirical predictions

Explicitly specifying the sound change and learning hypotheses above has raised an additional, untested empirical prediction: if low vowels' velar height is related to the typological difference in nasal contrasts, then low vowels should have a bigger *difference* in velar height between oral and nasal contexts. Henderson (1984) did not estimate the difference between nasal and oral contexts for low vs high vowels, only the *overall* difference in nasality between low vs high vowels, aggregated over contexts.To my knowledge, this prediction has not been systematically tested. Henderson (1984) and following replications found low vowel velum lowering in oral as well as nasal contexts, and while Henderson (1984) observed the difference between oral/nasal contexts might be smaller for low vowels, it was not estimated statistically. An additional prediction is that in production data, the shape of the distributions of [bVn] and [bVd] tokens should be less overlapping for low vowels.

Testing these articulatory and production predictions is beyond the scope of this dissertation (although relevant data may be available from Kunay et al. (2019)), but one of its contributions is to identify this testable prediction for future work, directing the investigation of the relationship between phonetics and typology by concretely specifying linking hypotheses about category learning.

Chapter 2: Testing the perceptual predictions of the height/length/nasality bias, MOG learning, and the typological nasal contrast asymmetry

If the typological asymmetry stems from the influence of vowel length on perception, then the simple infinite Dirichlet process MOG model predicts that the distribution of perceived nasality for [bVn] and [bVd] is less overlapping for long vowels, parallel to the predictions of the production bias in Chapter 1.

Although there has been some experimental work showing longer vowels are perceived as more nasal (Whalen & Beddor 1989; Hajek and Watson 1998), none has examined the effect's interaction with consonantal context. The effect of consonantal context by itself on perceived nasality has been investigated, with conflicting results (Kingston and Macmillan 1991; Macmillan and Kingston 1995 cf. Beddor & Krakow 1999; Zellou 2017). Additionally, many of these studies (Whalen & Beddor 1989; Hajek and Watson 1998; Beddor & Krakow 1999; Macmillan and Kingston 1995) estimate perceived nasality using English speakers' metalinguistic nasality judgements on a 1-5 scale that makes interpreting effect sizes difficult because it is unclear how individuals are using the scale and what cues they use for their explicit "nasal" judgements.

The goal of this chapter is to test the phonetic bias and MOG model's prediction of less distributional overlap in perceived nasality for longer vowels, using a discrimination task instead of metalinguistic identification. However, there are two challenges with this design: (1) discrimination tasks can only reveal the perceptual distances between stimuli, not their relative values of perceived vowel nasality. To

address (1), I pre-define a procedure for identifying the perceived nasal vowel dimension in multidimensional scaling output, by using stimuli with extreme acoustic nasality and orality as landmarks. (2) The model prediction is about exemplar distribution overlap rather than individual stimuli. To address (2), I focus on stimuli with nasality values that would be close to the region of overlap between oral and nasal vowels. If a perceptual bias reduces exemplar distribution overlap, it should increase the perceptual distance between oral-context and nasal-context vowels: given the same acoustic vowel nasality, vowels in a nasal context should be pushed to a greater perceived vowel nasality. Although this prediction and result would be surprising given compensation for coarticulation, there is evidence that sounds tend to perceptually assimilate to the following context (Rysling, Jesse, Kingston 2019), although this assimilation has not been tested for nasality.
The summary of the experimental design is given here:

- Research question #1: Does preceding a nasal consonant (e.g. “ban”) result in a vowel having greater perceived nasality?
    - Hypothesis 1a: Yes; this predicts that, for stimuli with acoustically identical vowels, the stimuli in a nasal context (e.g. “ban”) will have a greater degree of perceived nasality than stimuli in an oral context (e.g. “bad”)
    - Hypothesis 1b: No; this predicts that, for stimuli with acoustically identical vowels, the stimuli in a nasal context (e.g. “ban”) will have a lesser degree of perceived nasality than stimuli in an oral context (e.g. “bad”)
- Research question #2: If vowels in a nasal context are perceived as more nasal, is the effect greater for longer vowels?
    - Hypothesis 2a: Yes; this predicts that, for stimuli with acoustically the same amount of vowel nasalization (operationalized as the absolute duration spliced from a nasal vowel), the difference in perceived nasality between oral and nasal contexts will be greater for longer vowels
    - Hypothesis 2b: No; this predicts that, for stimuli with acoustically the same amount of vowel nasalization (operationalized as the absolute duration spliced from a nasal vowel), the difference in perceived nasality between oral and nasal contexts will not be greater for longer vowels

Participants will hear a pair of stimuli and rate their similarity on a scale (Wright 1977), which will provide estimates of perceptual distances. This method is used instead of a more typical discrimination paradigm, like 4IAX, because it allows more flexibility in the type and number of stimulus pairs used. The stimuli will consist of vowels with different acoustic levels of nasality, operationalized by the duration of the vowel that is spliced from a nasal vowel. Each vowel will occur in a stimulus with an oral context (b_d) and nasal context (b_n). The stimuli will include a completely oral vowel (henceforth labeled as V0), as well as an unambiguously nasal vowel (which will require a pilot experiment to determine which value of nasality results in consistently “same” ratings) of acoustic vowel nasality (labeled as V2), and an ambiguous degree of acoustic vowel nasality (labeled as V1). The stimuli will also vary on vowel duration, so the stimuli will be combinations of: level of acoustic vowel nasality (V0,V1,V2) x context (b_n or b_d) x vowel length (long or short, informed by the typical duration difference between high and low vowels). Because there are 12 stimuli (3 acoustic nasality levels x 2 contexts x 2 lengths), there will be 144 unique stimulus pairs (including swapped orders and same-stimulus pairs) for each participant to listen to, with 3 repetitions of each pair (= 432 trials per participant).

If the perceptual distance results roughly follow the triangle inequality, I will input them to metric MDS; otherwise, I will use nonmetric MDS. I will define the perceived vowel nasality dimension with this method: In the 2D MDS output, I will identify the midpoint between the oral context and nasal context V2 stimuli (i.e. the maximum acoustic nasality); this corresponds to high perceived vowel nasality. I will then identify the midpoint between the oral context and nasal context V0 stimuli (i.e. the minimum acoustic nasality); this corresponds to low perceived nasality. I assume the perceived nasality dimension corresponds to the line between these two midpoints. Macmillan and Kingston (1999) use a similar procedure to identify the angle between the perceived F1 and nasality dimensions. To get the perceived nasality values of the ambiguously nasal stimuli, I will project their 2D MDS locations onto the vowel nasality dimension, estimated as described above. I will evaluate Hypotheses 1a, 1b, 2a, 2b using the perceived nasality values of the stimuli with medium/ambiguous levels of vowel nasality (V1).

Chapter 3: Extended models of category learning

3.1 Size and implementation of phonetic bias

This chapter will further explore sound change mechanism hypotheses joint predictions with phonetic biases. Regardless of evidence for a perceptual bias found in Chapter 2, or further data on production bias in Chapter 3, this chapter asks the question: assuming there is some phonetic bias, what would it need to look like to affect the development of nasal vowels from coarticulation to allophony? While Soskuthy (2013) has significant modeling findings on the predictions of phonetic biases over time assuming clustering learners, the focus of this chapter is on the splitting of categories under an infinite Dirichlet Process Mixture of Gaussians learner. This chapter is also motivated by the findings of Smith et al. (2017), that the strength of a bias against patterns with variation has unintuitive typological predictions given a computational model of cultural evolution.

3.2 Time course of category splits given phonetic biases

This chapter will also explore the effect of several modeling assumptions in the sound change literature, namely the use of sub-categories (Soskuthy 2013; Gubian et al. 2023, and others), and assumptions about how changes are transmitted within infinite Mixture of Gaussians learners (across generations or between agents; suddenly or gradually).

While there's an existing category learning model that can represent phoneme splits (Gubian et al. 2023), this model has a different population/transmission structure than the one I hypothesize in Chapter 1: in Gubian et al. (2023), agents represent individuals of the same generation that interact through taking turns in production and perception, rather than transmission across generations. This hypothesized mechanism for phoneme split sound changes may make incorrect predictions for the development of phoneme splits, given evidence presented by Labov (1994): adults do not acquire phoneme splits through interaction with speakers that have the split, for example the the [a] split in Philadelphia and New York English. This inability for adults applies even when there is a social motivation: adult British English speakers with a [ʌ,ʊ] merger very rarely acquire the [ʌ,ʊ] split present in the prestige variety of Received Pronunciation. Thus, it appears they result from inter-generational transmission in first language acquisition rather than through the interaction of adult speakers. It is not clear yet whether the two definitions of sound change model, agent-based interaction versus transmission across parent/child generations, make substantially different predictions about phoneme splits; Soskuthy (2013) demonstrates that some exemplar models make the same predictions for phonetic biases regardless of population structure.

3.2 Phoneme split mechanism: drift problem

A more general problem, regardless of phonetic biases, is how a category split would work gradually in a MOG sound change model as opposed to mislearning the number of categories in a single generation. Consider (1) an initial two-category system with coarticulation affecting vowels in a nasal context, which shifts into (2), a three-category system that splits vowel categories based on nasality.

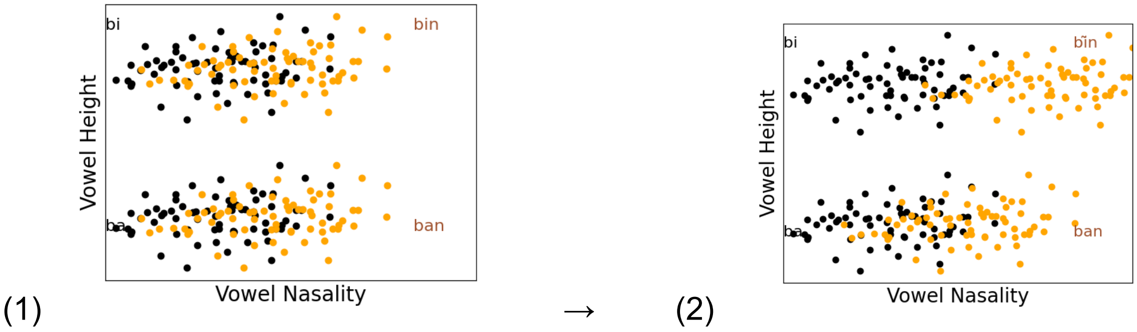

Figure 3.2.1. Illustration of two possible category representations of an individual, one at an earlier stage time and one at a later time. The first learner has one category for each height, with nasal-context tokens occurring with slightly higher nasality. The second learner has for high vowels two categories, one nasal and one oral.

The reason why a MOG sound change model wouldn't have a gradual shift from (1) to (2), with the oral and nasal vowel categories slowly drifting apart, is that a child learner of (1) can either learn a single category for bi/bin, with a single mean, or learns two separate ones, like in (2). There is no representation for any stages in-between, where the vowels in bi/bin represent a single category but the productions for bin are intentionally more nasal (or in other terms, sampled from a more nasal distribution) – MOG representations mean there's only one mean for each category.

Dillon et al (2013)'s model has something closer to representing this medial, "drifting" stage, because their MOG categories' means can be shifted dependent on context; the size and direction of the shift is continuous, which could allow for a gradual change over many generations. However, their model shifts all categories in the same amount and direction, so it cannot represent cases with nasal allophony for one category but not another, e.g. nasal allophony for low vowels but not high vowels. This shared shift allows the model to reduce the size of its grammar (e.g. two means and one shift - 3 parameters - instead of four means - 4 parameters); if the model were allowed to learn a separate shift for each category, there would be no difference from the model's point of view between shifting and learning separate categories (e.g. two means and two shifts - 4 parameters - versus four means - still 4 parameters).

Few existing sound change models focus on category splits; most focus on shifts, mergers, and contrast maintenance (e.g. Blevins & Wedel 2009, Harrington 2017; Kapatsinski 2021). The models in Stevens et al. (2019) and Gubian et al. (2023) models category splits by having agents checking their categories every few iterations and splitting any categories that would result in a significant improvement in data likelihood. It's unclear in this model whether there are "drifting" stages or whether the category split is modeled as individual variation, with some agents who haven't split and others who have, and the number of split-category agents growing through time. This model might be able to capture a gradual "drift" because although the categories are still Gaussian with a single mean and variance, there's an extra word layer of abstraction, where there is a distribution for each word inside of the distribution for the sound category (cf. Feldman et al. (2013)'s joint word-phone learning model). There's also the empirical question of how category splits actually occur, through a gradual drift or from individual variation (which has been suggested as an important factor in sound change, e.g. Stevens & Harrington 2014).

3.3 Category representations and inference problem

This section may not be feasible within my timeline, but I would like to explore the representational and task assumptions of the learning model.

Beddor (2009) proposes that vowel nasalization arises when nasalization is misparsed as a feature of the vowel as opposed to the nasal consonant, with a perceptual tradeoff between nasal consonant length and vowel nasality. Is this misparsing more likely for vowels that tend to be produced with a larger velar opening or for longer vowels? Beddor (2009) discusses this misparsing in terms of gestural representations. However, as there has been no work on developing a formal, language-specific category (gesture constellation) learning algorithm for gesture-based phonology, this might be beyond the scope of this dissertation.

How is this perceptual misparsing implemented in perception and category learning? Kirby (2014) takes a step toward implementing misparsing in sound change with a Mixture of Gaussians model of /CrV/ sequences becoming /ChV/ sequences: with some probability for each input /CrV/ token, parts of the trill implementation of /r/ are misperceived as aspiration, with the duration re-assigned from /r/ to the the preceding consonant's VOT. However, in the case of vowel nasalization, the ambiguity is not in the cue (trilling versus aspiration) but in the underlying segment it is associated with (nasalization coming from a consonant versus vowel).

I could design a Bayesian learner that infers the intended source (vowel category or nasal consonant) of observed nasality, and examine if it has any structural biases that favor assigning nasality as an intended realization of a vowel if it generally is produced with more nasality.

An additional representational question worth exploring is the shape of the perceptual space for nasality: is there a threshold of acoustic nasality for a sound to be perceived as nasal? How does including a threshold affect the predictions of a Mixture of Gaussians or Bayesian category learner?

Chapter 4: Application to voicing contrast case study

According to the P-Map hypothesis, there's a universal hierarchy of contexts where voicing contrasts are licensed (Steriade 1997; Yu 2004). A constraint penalizing voicing contrasts in a context where a difference in voicing is less perceptible (e.g. word finally) is universally ranked higher than a constraint penalizing voicing contrasts in a more perceptible context (e.g. between a vowel and a sonorant). The contexts where a language has a voicing contrast is determined by the ranking of Preserve (voice), a faithfulness constraint, relative to the universally ranked constraints on voicing contrasts. Under this system of constraints, the P-Map predicts that a final voicing contrast implies a voicing contrast between sonorants; in other words, if a language has a voicing contrast word finally (a less perceptible context), it will also have a voicing contrast between sonorants (a more perceptible context).

According to the Evolutionary Phonology hypothesis (Blevins 2004), certain contrasts are more common in certain contexts not because of a cognitive restriction on possible grammars, like the P-Map, but instead because of the relative frequencies of the sound changes that result in these contrast systems. For example, final devoicing is a common sound change that can neutralize word-final voice contrasts, so Evolutionary Phonology would predict voice contrasts to be less common word-finally than in other positions that do not result from a common sound change.

The predictions of Evolutionary Phonology and the P-Map may diverge in the case of intervocalic voicing. Intervocalic stop voicing appears to be a common sound change, but languages with intervocalic stop voicing neutralization are rare (cf. Katz 2016).

Suppose there is a parent language with a plural suffix -a and no voicing alternations, e.g. with two lexical items
rat ~ rata
rad ~ rada

Now suppose this language undergoes intervocalic stop voicing, a putatively common sound change.
*t > d / V_V

The two lexical items from our toy language are now realized as:
rat ~ rada
rad ~ rada
In this resulting language, there is now a voicing contrast word-finally but not between sonorants, which is putatively rare or unattested. The P-Map hypothesis correctly predicts this system to be unattested, or at least unproductive, as the hypothesized universal constraint system for voicing contrasts requires any language with a word-final voicing contrast to also have a between-sonorant voicing contrast. Evolutionary Phonology, however, does not predict this system to be rare, as the sound change that could cause it (*t > d / V_V) is common.

However, the exact mechanisms of final devoicing (merger of stop categories word-finally) and the development of voicing distinctions (split of stop categories in other contexts) have not been well-defined. Because they involve category splits and mergers, the category learning and sound change hypotheses developed in this dissertation may be applicable to them and generate predictions to distinguish the P-Map and Evolutionary Phonology accounts.

Because I am still developing the category learning and sound change hypotheses, it is not clear yet what predictions each one makes for the mergers and splits of stop voicing categories in different contexts.