Chapter 4

Intraspeaker Variation in æ

An investigation of the mechanism of phonological change as driven by individual speakers relies on an in-depth analysis of the production of individual speakers. In this chapter, I present an analysis of transitional cohort speakers, finding that their data is most consistent with change via competing grammars.

Despite the Philadelphia Neighborhood Corpus providing data spanning the entirety of this phonological change and the Influence of Higher Education on Local Phonology corpus providing key data into the community-level fragmentation of this change, the amount of per-speaker data provided in these corpora do not allow for a robust analysis of change. Here, I create an additional corpus of speech designed specifically to target transitional cohort speakers and to obtain enough test tokens of /æ/ from each speaker to identify the mechanism of phonological change. This corpus, which I refer to as Investigating the Mechanism of Phonological Change (IMPC), is described in §4.3. My method for analyzing individual tokens is outlined in §4.1, and the predictions that each mechanism of change makes for the production of /æ/ by transitional cohort speakers are discussed in §4.2. In §4.5, I analyze all speakers in the IHELP and IMPC data sets that produce enough data to bear on the mechanism of phonological change.

4.1 Analysis of Individual Tokens

To analyze the mechanism of phonological change through the production of individual speakers, it will be necessary to analyze each test token as having been produced by either PHL, NAS, or some intermediate phonetic incrementation of these two systems. In most cases of phonological change, particularly for changes involving phonological mergers or splits, classifying individual tokens as having been produced by the old or the new phonological system is almost always impractical, given the overlapping distributions of tokens in phonetic space and the number of observations collected for typical sociolinguistic data. The allophonic restructuring from PHL to NAS, however, provides a rare opportunity to classify each individual observation according to which underlying system it adheres to. Here, I use the structural similarities and differences between PHL and NAS to my advantage. For PHL, as for NAS, a token of /æ/ preceding a tautosyllabic front nasal (as in *hand* or *ham*) will be produced as lax. Likewise, there are a large number of words that fall into the elsewhere condition for both systems, producing lax /æ/ in words like *cat* and *dad*.

The shared conditioning between PHL and NAS means that we do not need to know whether a speaker has the PHL system or the NAS system in order to characterize that speaker's tense and lax acoustic targets: their tense target will be in the phonetic space of HAND words and their lax target will be in the phonetic space of CAT words. This information about a speaker's phonetic targets can then be used to determine whether each test token aligns best with that speaker's tense target or their lax target.

4.1.1 Classification Methods for Test Tokens

The problem of classifying test tokens for tense, lax, or intermediate realization is not trivial. To determine the optimal classification method, I test different classification methods to actual data from a PHL speaker and a NAS speaker, as well as to simulated data (described in detail in Appendix B), to determine which classification system produces interpretable results for transitional cohort speakers. The classification methods attempted include K-means cluster analysis and Hierarchical cluster analysis run on the F1 and F2 values for tokens as well as these methods run on a Principle Components Analysis (PCA) and a t-Distributed Stochastic Neighbor Embedding (t-SNE) transformation of the data. For the purposes of this dissertation, a glm classifier provided the best fit to the data, and is the method used and described here.

A generalized linear model (glm) is a family of linear regression models, which can be turned into a classification method. As a classification method, a glm model is first fit to training data, which provides coefficients for each term in the model. These coefficients are then used to predict the outcome of test data. Typically, this method is used as a basis for machine-learning: human coders code a random subset of a data set, and the resulting glm model for that training data is then fit to the rest of the test data. As a method for classifying /æ/ production for transitional cohort speakers, this must be slightly modified. We cannot take a random subset of data, precisely because we can not determine *a priori* whether a speaker's test tokens are tense or lax. However, because of the overlapping conditioning factors between PHL and NAS, we can determine the phonetic target of a speaker's HAND and CAT class tokens. Here, I use these tokens as training data for a glm classifier for each speaker, which is then fit to test data to predict whether each test token was produced as phonetically tense or lax.

4.1.2 Applying the Glm Classifier to Speaker Data

The first step in using a glm classifier is to split a speaker's data up into training tokens and test tokens. Here, we use each speaker's HAND class tokens as training tokens for the tense phonetic target and CAT class tokens as training tokens for the lax phonetic target. Figure 4.1 shows the training tokens for Bobby Marx, a Philadelphian born in 1967 whose data is part of the IHELP corpus. 95% confidence ellipses are plotted around Bobby's HAND class tokens (solid line) and his CAT class tokens (dashed line), to show the acoustic characteristics of his tense and lax targets, respectively.

For each speaker, I use these training tokens to train a glm classifier on the acoustic parameters of that speaker's tense and lax targets, using F1 measurement, F2 measurement, duration, and syllable stress as independent variables, shown in (14). The resulting coefficients of this classifier were then used to predict the probability of tense or lax realization for the remaining test /æ/ tokens using the `predict()` function, shown in (15).

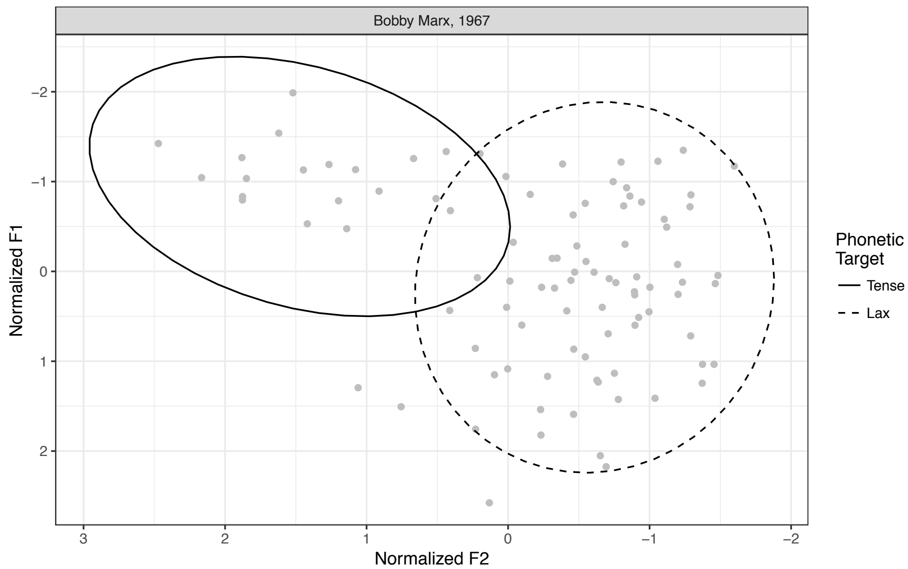


Figure 4.1: Bobby Marx training data.

(14) `predmod <- glm(tense ~ F1*F2*F3*duration*stress)`

(15) `testdata$tenseProb <- predict(predmod)`

Once test tokens have been classified as either tense or lax, each token can then be classified as having been produced by PHL or NAS. Any LAUGH or MAD token that has been classified as tense by the classifier is consistent with PHL but not NAS; likewise, a MANAGE or HANG class token that is classified as tense is consistent with NAS but not PHL. In Figure 4.2, we can see the results of the classifier. Training tokens are again plotted in gray, with 95% confidence ellipses drawn around the tense target (as identified through HAND class tokens) and lax target (as identified through CAT class tokens). Test tokens are plotted over the training data, with tokens classified as PHL in orange and NAS in green. We can see in Figure 4.2 that Bobby overwhelmingly produces tokens consistent with PHL; given his demographic data as a Philadelphian born before 1983, we expect to find predominately PHL data. However, it is also clear that there are a few tokens that are selected by the glm classifier as consistent with NAS. Given that Bobby's overhelming pattern is PHL, I term these tokens *incongruent tokens*, as they are incongruent with his dominant system.

The distribution, number, and lexical identity of these incongruent tokens are an important aspect of identifying whether a speaker's production matches PHL, NAS, or an intermediate system. I come back to this in §4.5.

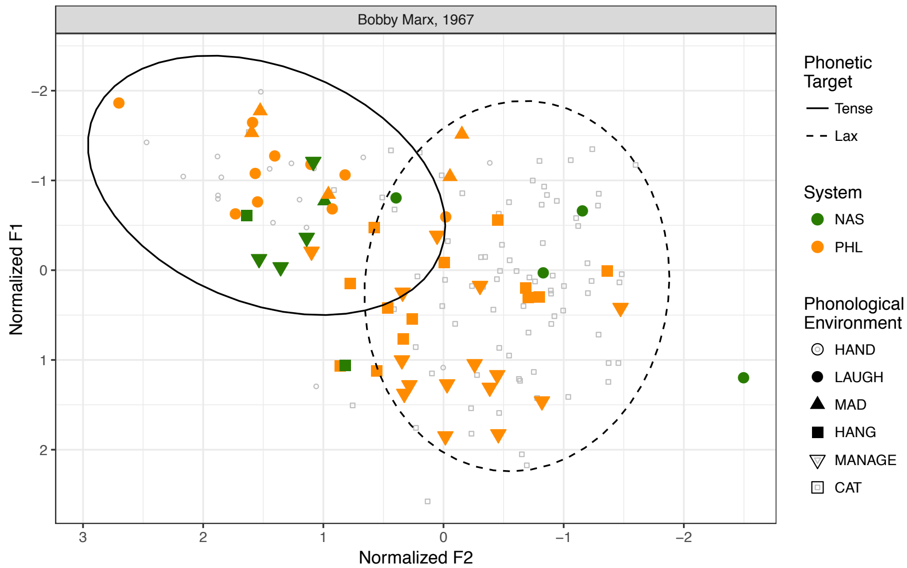


Figure 4.2: Bobby Marx test data.

There is a final point to make about the use of the glm classification system. While the glm classifier produces a probability value for each test token between 0 (lax) and 1 (tense), the break point between which tokens are classified as tense or lax is not 0.5. This is particularly true, given the phonological conditioning of PHL tense and lax tokens. All prenasal tokens in the training data are tense, and all tense training data are prenasal, producing a bias towards classifying prenasal test tokens as tense. Adding to this bias is the acoustic output of PHL tense tokens: Kroch (1996) found, for example, that prenasal tense tokens are acoustically higher and fronter than the LAUGH and MAD class words[7]. To determine the most accurate cutoff thresholds for tense classification, I use the productions of traditional PHL and NAS speakers to obtain a probability threshold that maximizes the accuracy for both types of speakers. This results in a cutoff of 0.22 for prenasal tokens (above which a token will be classified as tense),and a threshold of 0.14 for non-prenasal

[7]While prenasal tokens are realized as acoustically more tense than the rest of the traditional tense class, an ultrasound study (Mielke et al., 2017) finds all tense tokens of the traditional PHL system to be articulatorily identical with regards to tongue position.

tokens.

After being classified as tense or lax, each test token was then categorized as either a PHL token or a NAS token, according to which system it was consistent with. Tokens categorized as PHL are represented in orange, and tokens categorized as NAS are represented in green.

4.2 Profiles of Each Mechanism of Change

Because of the complex set of facts surrounding PHL, it is useful to first run through the predictions that each mechanism of phonological change make. In what follows, I present simulated data to highlight the predicted profile for each mechanism of change. For each simulated change, I create a hypothetical PHL speaker and a hypothetical NAS speaker, then create transitional generation data for three intermediate steps based on the assumptions from each mechanism of phonological change.

The simulated PHL data is generated using F1, F2, covariance matrices, and token count values drawn from an actual PHL speaker (Mary C., whose production of PHL represents a prototypical PHL speaker and who produces one of the highest token counts of /æ/ in the PNC, with $N = 1456$). Simulated NAS data is generated using these same values from Cara G., who is the speaker in the IHELP corpus with the highest token count ($N = 825$). Simulating the productions of PHL and NAS allows me to set the seed for each simulation, resulting in pseudorandom simulated tokens. Setting the seed to the same number for each of these plots enables us to see that any changes between plots is due to differences in the underlying means and covariances of the plots, rather than due to random noise in generated data. Given the PHL speaker and NAS speaker, I then generate transitional generation data following the assumptions of each mechanism of change, which are described in detail below.

Figure 4.3 compares Mary C's actual production data (left) with the simulated plot of her production data (right). In both facets, a 95% confidence ellipse is drawn around the HAND class (solid line) as well as around the CAT class (dotted line), to give a visual representation of her tense and lax phonetic targets. Simulated data was created using the mvtnorm package in R, with F1 and F2 means for each conditioning factor drawn from Mary's data and covariance matrices for F1 and F2 produced with Mary's actual covariance measures. The simulated data contains the same token count for each conditioning factor as Mary's actual data, so that the simulated data contains an accurate snapshot of the relative proportion of /æ/ tokens within each conditioning factor. Mary's simulated data is shown in the right panel.

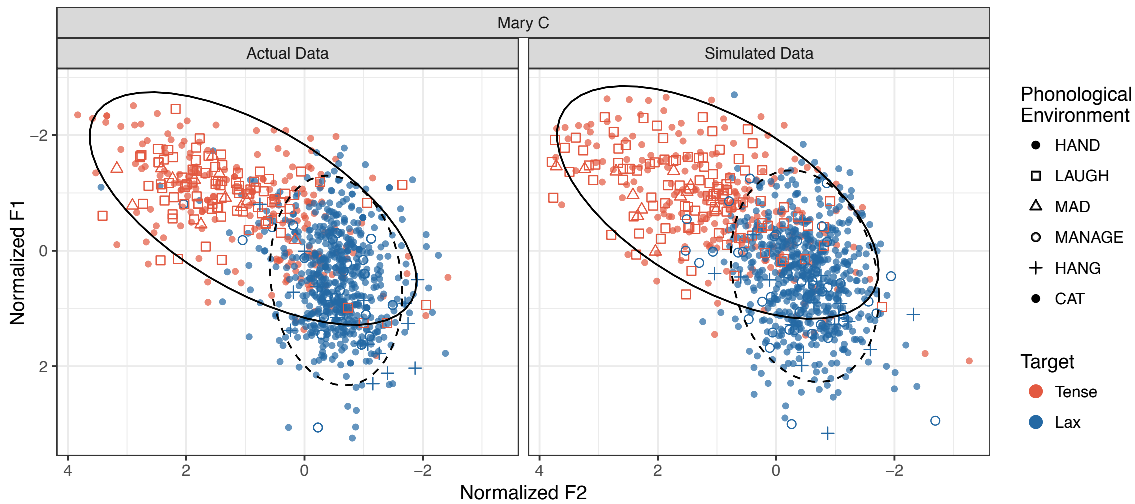


Figure 4.3: Mary C. real data (left) and simulated data (right).

Figure 4.4 displays a similar output for Cara G, with the F1 and F2 means drawn from Cara's actual production of /æ/ in each conditioning factor, and covariance matrices calculated separately for each conditioning factor. The means and covariance for the simulated data matches Cara's production; however, here we have drawn the *N* values per conditioning factor from Mary's production, so that our simulated speakers are maximally comparable. In other words, the right hand panel depicts the output we would expect if Cara had produced 1456 /æ/ tokens instead of 825.

I use these simulated plots of PHL and NAS so that we may produce a 5-step continuum between PHL and NAS based on the specific predictions from each mechanism of phonological change. Each plot is generated from a pseudorandom gaussian distribution. Using `setseed()` ensures that each plot is generated from the same seed, resulting in psuedorandom rather than fully random data. This allows us to reproduce each plot, changing only the F1, F2, and covariance parameters as predicted by each mechanism of change. In other words, any differences in the position of a particular token between two plots is due to an actual difference in the theoretical predictions

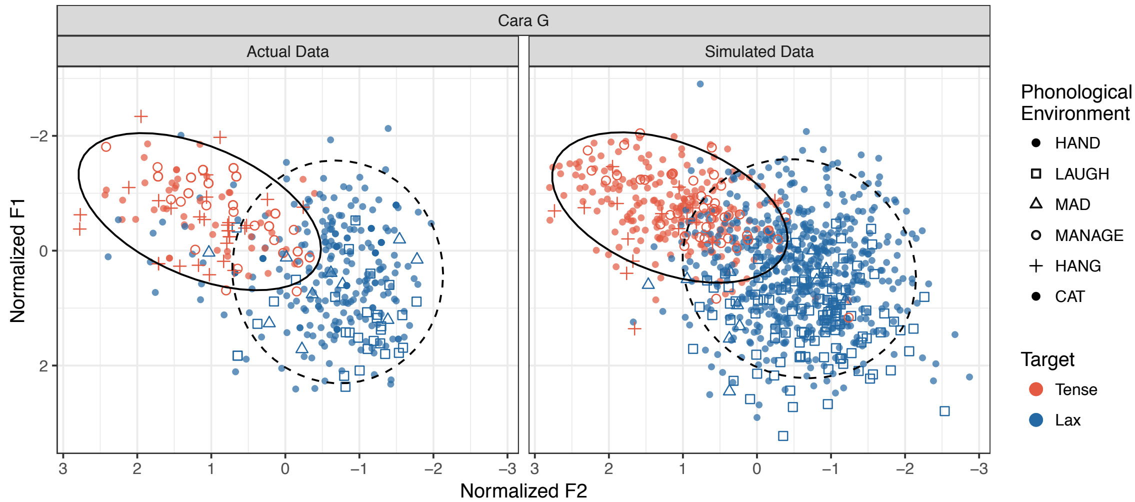


Figure 4.4: Cara Grant actual data (left) and simulated data (right).

rather than to noise in the data. Each step in the 5-step continuum is labelled as a different cohort for the sake of temporal exposition. The actual time difference between "Cohort 1" and "Cohort 2" may only be a short number of years; the main point is that each panel in the following plots represents one speaker who is slightly more advanced in the change from PHL to NAS than the previous panel.

For each simulated speaker, I run the simulated data through our tenseness glm classifier, based on the F1 and F2 values for tense and lax for that simulated speaker's HAND and CAT classes. Each simulated token is then classified as either consistent with PHL (orange) or NAS (green), which provides a profile of what our expected outputs from the transitional cohort speakers will be.

4.2.1 Phonetic Incrementation

There are a few possible profiles for phonetic incrementation to follow in this case, and these depend on the unit that is being incremented.

Tense and Lax Incrementation

The naïve hypothesis for PHL becoming NAS through phonetic incrementation is that phonetic incrementation affects both the tense allophone of /æ/ and the lax allophone simultaneously. The idea here is that due to some unspecified combination of production and perception errors, all conditioning factors contributing to the tense allophone of /æ/ become phonetically laxer while all conditioning factors of the lax allophone becomes phonetically tenser. In the middle of the change, we would expect both allophones to be produced in the same phonetic space, in between canonical PHL lax and canonical PHL tense and appearing merged in phonetic space. After this middle stage of the change, we would expect to see the hint of allophonic restructuring, with the conditioning factors contributing the tense allophone of NAS becoming phonetically raised while the conditioning factors contributing to the lax allophone of NAS become phonetically laxed, leading to a final stage where the production of /æ/ matches our prototypical NAS speaker. The simulated data is created using a 5-step linear interpolation between the F1 and F2 means and covariance matrices of the simulated PHL and NAS data.

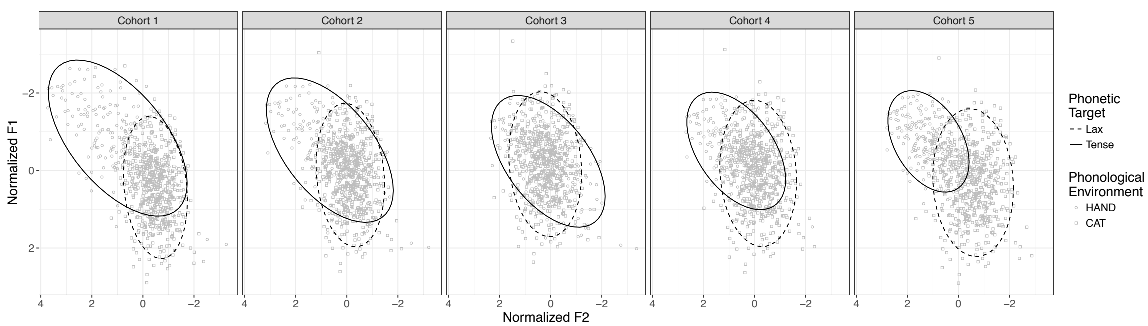


Figure 4.5: Profile of phonological change for HAND and CAT classes under phonetic incrementation of all conditions.

For all the changes that follow, we produce specific predictions about the behavior of HAND and CAT as well as the four conditions that differ between PHL and NAS. For clarity, here we present first the predictions about the shared conditioning factors first, before overlaying the predictions about the test conditions. Figure 4.5 displays the predicted acoustic outputs of HAND and CAT given full phonetic incrementation. As we can see, the two acoustic targets drift together in pho-

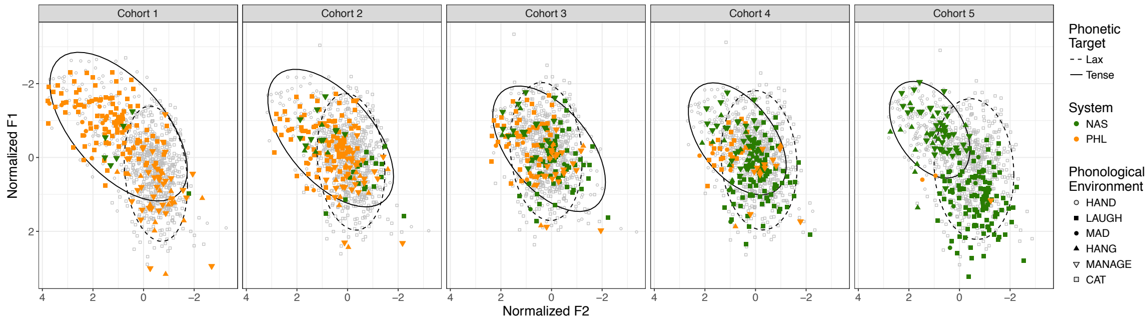


Figure 4.6: Profile of phonological change for test classes under phonetic incrementation of all conditions.

netic space, completely overlapping in Cohort 3, then drift apart again. The test conditions follow suit (4.6); each of the four test conditioning factors (LAUGH, MAD, HANG, MANAGE) increment towards a central position in Cohort 3, then continue on their merry restructuring way to produce full NAS by Cohort 5. I note briefly that like the actual data discussed above, these simulated productions of PHL and NAS contain some test tokens that are classified as incongruent with the rest of the tokens. In analyzing actual speaker data, it is a close analysis of test tokens such as these, that suggest underlying systemic variation, that will provide an account of the mechanism of phonological change.

The main identifying characteristics of phonological change through this type of phonetic incrementation are in the unimodal distribution of all conditioning factors in Cohort 3; the training tokens are merged in acoustic space, as are the test tokens.

Tense Allophone Incrementation

Given the sociolinguistic facts reported for PHL however, we expect phonetic incrementation as shown in §4.2.1 to be unlikely. Most relevantly, Labov (2001) finds Philadelphians who produce PHL to negatively rate only the tense tokens of PHL. This negative evaluation produces a social motivation for phonetically laxing the tense allophone but not the lax allophone. This prediction of phonological change is very similar to the prediction described above; here the only difference is that the lax allophone remains in its lax position throughout the change, while the tense target shifts down to join it in the lax target before moving back to the NAS tense position (Figure 4.7). Here, the transitional cohort values are created first using a 3-step linear interpolation of of F1, F2, and covariance matrices between PHL and Cohort 3 and then a 3-step interpolation between Cohort 3 and NAS. Cohort 3 was created using F1 and F2 for the mean of the lax test class, so Cohorts 1-3 represent a gradual shift of all tense PHL tokens to the CAT target while Cohorts 3-5 represent the gradual shift of tense NAS tokens from the CAT target to the NAS tense target. As with the previous simulation, covariance matrices are a 5-step interpolation between the covariance matrix of PHL and the covariance matrix of NAS.

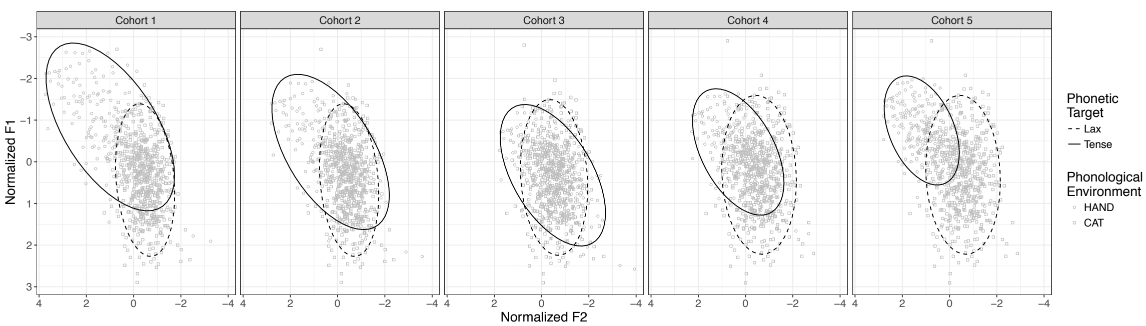


Figure 4.7: Profile of phonological change for HAND and CAT classes under phonetic incrementation of tense allophone.

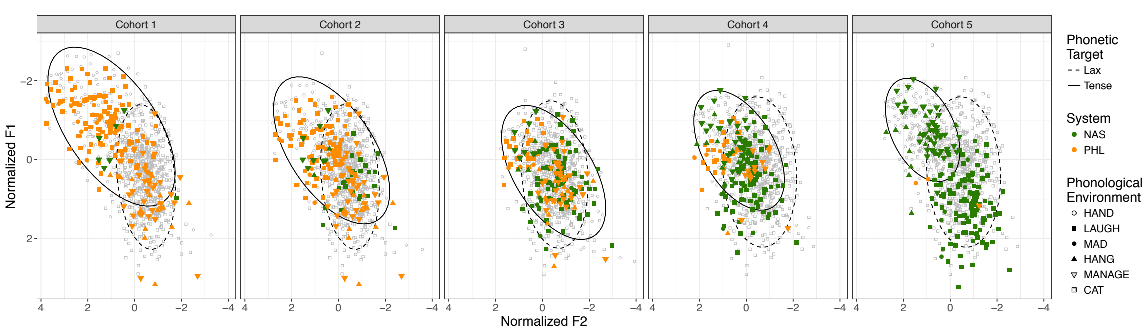


Figure 4.8: Profile of phonological change for test classes under phonetic incrementation of tense allophone.

The characteristics of this change are nearly identical to those laid out above: we find a unimodal distribution of both the training tokens and test tokens in Cohort 3; the only difference here is that we find this unimodal distribution occurring in the acoustic space of speakers' lax targets.

Test Conditioning Factors

The unimodality in all tokens predicted by the phonetic incrementation profiles above are easy to identify as change via phonetic incrementation. Unfortunately, phonetic incrementation could also take on a less clear path. Because phonetic incrementation does not rely on any abstract processes, this mechanism allows for any group of tokens to shift or remain unshifted. This is specifically in contrast to a phonologically based theory of change, in which the target of any phonetic movement is predicted to be a cohesive phonological unit or phonological feature. In other words, there is nothing baked into the theory of phonetic incrementation that predicts that all tokens of an allophone or even all tokens of a phonological conditioning factor will necessarily undergo the same set of errors in production and perception. It is possible, then, that this change from PHL to NAS is the result of phonetic incrementation of only a subset of tokens. Here, we present the most likely version of this, in which the conditioning factors that differ between PHL and NAS phonetically increment but the shared conditioning factors remain stable. In this simulation, the HAND and CAT classes are produced as a 5-step linear interpolation between PHL and NAS F1, F2, and covariance matrices. The test conditioning factors are produced in a 3-step interpolation between PHL and Cohort 3, then a 3-step interpolation between Cohort 3 and NAS. Cohort 3 is produced using the mean F1 and F2 of the HAND and CAT classes and the mean covariance matrices between PHL and NAS for each conditioning factor.

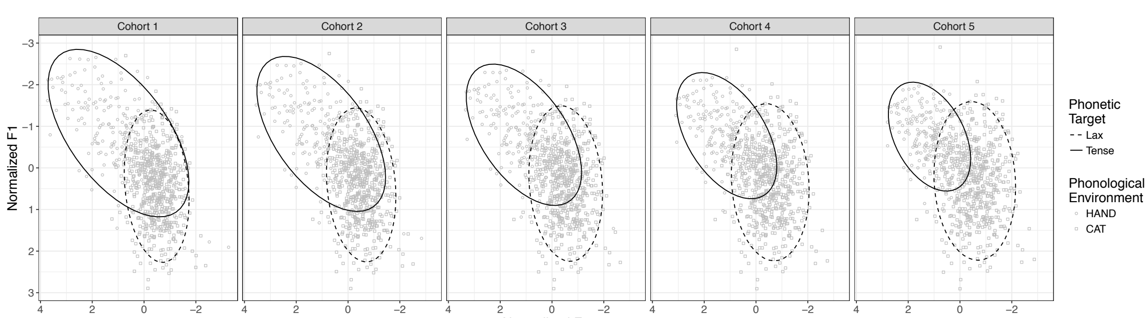


Figure 4.9: Profile of phonological change for HAND and CAT classes under phonetic incrementation of test conditions only.

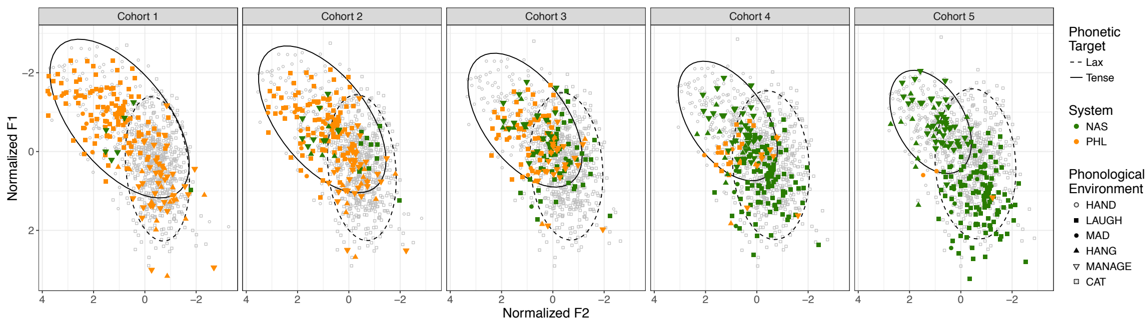


Figure 4.10: Profile of phonological change for test classes under phonetic incrementation of test conditions only.

Figures 4.9 and 4.10 show the predicted outputs for the training and test tokens, respectively. Unlike the previous predictions for change via phonetic incrementation, this type of change predicts that the shared conditioning factors will remain as a tense and lax target; we expect to see the training tokens exhibit bimodal distribution. The test tokens, on the other hand, still go through a period of unimodality in Cohort 3, where they are produced in the intermediate space between the tense target and the the lax target.

As we will see for the predictions made by spontaneous phonologization and grammar competition, it is this unimodality of the test tokens that provides the strongest cue for change via phonetic incrementation.

4.2.2 Spontaneous Phonologization

If the change from PHL to NAS is driven by the transmission mechanism of spontaneous phonologization, this predicts that individual speakers will posit a single allophonic system (either PHL or NAS), and stick to this system in their production. The community-level change, then, will be driven by an increasing number of speakers positing NAS in each successive cohort. This is represented in Figure 4.11, which depicts a representation of four speakers in each cohort. In the first cohort of speakers, representing traditional PHL in the community before any posited change to NAS, every speaker posits and produces PHL. By Cohort 2, one speaker out of four has posited NAS. This increases until Cohort 5, in which every speaker has posited and is producing NAS.

If the change on the community level has been driven by spontaneous phonologization, each

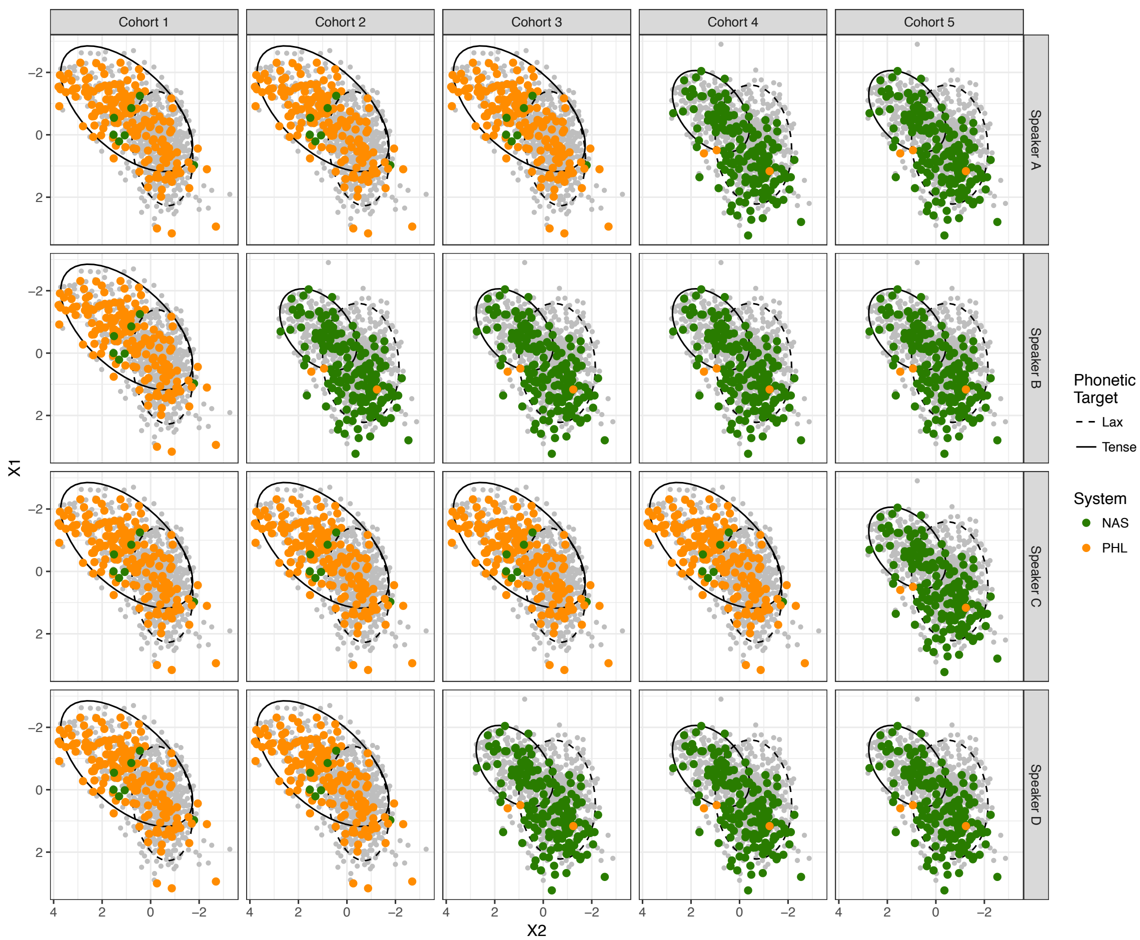


Figure 4.11: Profile of phonological change for change via spontaneous phonologization.

individual speaker should produce strictly PHL or strictly NAS outputs. This mechanism of phonological change requires an analysis of the community as a whole: given a single individual, it will not be possible to determine whether that speaker displays change via spontaneous phonologization or whether that speaker simply is drawn from a subset of the community that has completed the change to NAS or not yet begun the change to NAS. Determining whether change is occurring via spontaneous phonologization will require sociolinguistic data about that speaker's educational peers, as discussed in Chapter 2.

4.2.3 Intraspeaker Grammar Competition

If the change from PHL to NAS is driven through the mechanism of intraspeaker grammar competition, we expect to see a bimodal distribution of all test conditioning factors. This theory of change states that allophonic /æ/ system is a parameter in speakers' grammars, which in this period of change varies between the PHL variant and the NAS variant within a single speaker. In other words, allophonic systems as a whole would act as an abstract level of sociolinguistic variable, with speakers using some proportion of each system. This predicts that speakers in the beginning of the change are using mostly PHL tokens while speakers in the end of the change are using mostly NAS tokens. Here we present only the prediction plot of all tokens, since there is no main difference between the test tokens and the training tokens. Figure 4.12 displays the simulation results for all five steps of this change, beginning with 100% PHL tokens and ending with 100% NAS tokens. Cohort 2 is comprised of 25% NAS tokens, Cohort 3 is comprised of 50% NAS tokens, and Cohort 4 is comprised of 75% NAS tokens.

The predictions for change through grammar competition are clear: in all cases, we expect to see a distinct tense and lax acoustic target from the shared HAND and CAT classes. Test tokens are produced well within the tense and lax targets, with a bimodal distribution, and we expect to see variation between PHL and NAS at roughly equal rates across all test conditions.

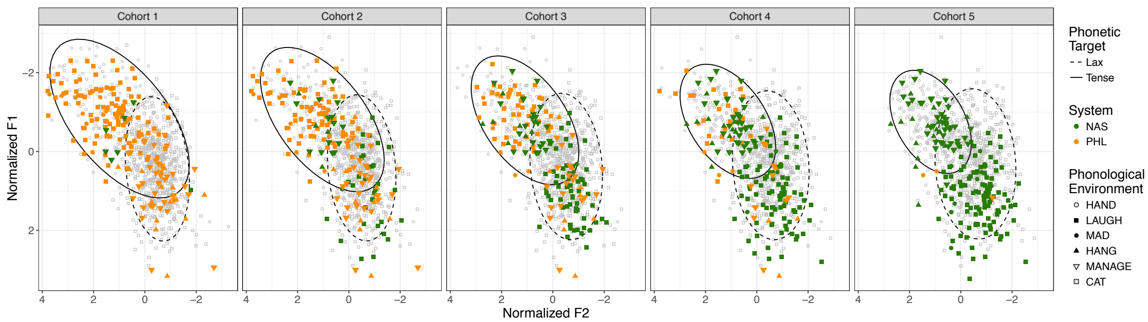


Figure 4.12: Profile of phonological change for change via competing grammars.

4.2.4 Summary of Predictions

Sections 4.2.1 through 4.2.3 provide an outline for predicted outputs based on each mechanism of phonological change. Here, the most crucial point of comparison is between the expected outputs for Cohort 3 from each of the theories of phonological change, since that is where we see the biggest differentiation between the theories. It is the profiles of Cohort 3 simulations that we are particularly looking for in our actual data. In what follows, I will use these profiles of predicted outputs to best characterize transitional generation speakers' productions. If we find speakers producing outputs that match one of the profiles of phonological change, this will serve as evidence that that particular mechanism of phonological change is at play.

4.3 Investigating the Mechanism of Phonological Change Corpus

The data that would bear on our particular question is relatively rare in frequency, both in terms of transitional cohort speakers and the number of tokens each speaker produces. I demonstrated in Chapter 2 that even amongst a single age range in Philadelphia, it is only within particular subsets of the population that PHL and NAS are currently vying for dominance. NAS has won out in elite non-Catholic high schools, and PHL still has a stronghold in the non-elite Catholic high schools. The social networks in which we expect the highest likelihood of mixed system speakers is in the graduates of non-elite non-Catholic high schools and of elite Catholic high schools. In addition to social networks resulting in only a subset of the population likely to acquire competing /æ/ systems, it is also the case that the specific test tokens of /æ/ that would disambiguate between PHL and NAS are rare in conversation. Within the IHELP data, for example, /æ/ tokens comprise only 14% of a speaker's data (with an average of 579 /æ/ tokens and 81 test condition /æ/ tokens per speaker over a one hour sociolinguistic interview).

In Chapter 2, I took a wide sampling approach in an effort to more fully describe the community-level pattern of this change. Here I must take a more targeted approach, with the goal of being able to analyze variation in speakers' test tokens in a way that will determine whether this change is occurring through phonetic incrementation, instant phonologization, or grammar competition. Drawing from the results in Chapter 2, I focus here on the recent graduates of Catholic schools, which is the population currently in flux with regards to this change and therefore the most likely to be transitional cohort speakers. In addition to targeting the population of speakers most likely to be transitional cohort speakers, it is also necessary to increase the number of test tokens obtained per speaker so that I may distinguish between surface-level variation that is driven by grammar competition and surface-level variation driven by another factor such as phonetic incrementation or lexical diffusion. The methods used to target transitional cohort speakers as well as to increase the yield of test tokens are described in some detail below. The data collected under the methods highlighted below is compiled into a single corpus Investigating the Mechanism of Phonological Change (IMPC).

4.3.1 Targeting Transitional Cohort Participants

The IHELP data resulted in some data that suggested that speakers were variable between PHL and NAS; however, there simply were not enough tokens from most of these speakers to rule out any proposed mechanism of phonological change. Here, I target those individuals who are the most likely to be transitional generation speakers. I do this in two ways. First, every participant from the IHELP database whose data suggested variation between PHL and NAS were invited to participate in this as a follow-up study; this resulted in four participants. Second, I extend my reach by targeting Catholic school graduates who were born after 1983, using my existing social networks in Philadelphia. This resulted in sixteen additional participants.

Participants obtained through social networks were first given a screening test, which consisted of a 10 minute Semantic Differential task (see Table 4.1 for the exact questions asked during screening) conducted over the telephone. Participants' productions of each test item were auditorily coded, and any participants who were found to produce variation were invited to participate in a full session. All participants were paid $30 per session they participated in.

4.3.2 Increasing Test Tokens through Interview Methods

Topic Directed Conversations

The first requirement for data collection is that I obtain over an hour of speech per speaker. My aim is to collect at least 10 tokens per test conditioning factor; judging from the rate of test tokens found in the IHELP corpus and the PNC, a classic sociolinguistic interview (Labov, 1984) would need to be roughly two hours in length. To increase the number of test tokens per hour of speech, I introduce the method of *Topic-Directed Conversations*. In this method, participants come into a quiet recording space with a friend, and are recorded as a dyad having a conversation. Following Boyd et al. (2015), the researcher leaves the room and allows the two participants to interact in a naturalistic way. One potential pitfall of using a dyad recording method instead of a traditional sociolinguistic interview is in the expected proportions of participant speech: while a sociolinguistic interviewer is trained to have the interviewee speak most of the time, a more natural conversation between two participants will result in each participant speaking roughly 50% of the time. I find, however, that volume of per-participant speech with participants each speaking roughly 50% of the time in a 1.5 hour Topic-Directed Conversation exceeds the average volume of a participant speaking roughly 80% of the time in a 1 hour sociolinguistic interview (avg. 4855 words vs. 2751 words).

In addition to enabling me to obtain naturalistic data from two participants at once, Topic-Directed Conversations also provide a more important benefit in that they direct participants toward topics with a high likelihood of producing relevant test tokens. For test tokens of /æ/, conversational prompts included the questions in (16)–(18).

(16) When is the **last** time you got really **mad**? Have you two ever gotten into a fight with each other? What do you do when you're **angry**?

(17) What about the **last** time you were embarrassed? Do you remember a time that one of your friends did something really embarrassing?

(18) When's the **last** time that you remember feeling scared? Is there anything that makes you feel like you're going to **panic**?

In (16)–(18), I've bolded the test tokens inherent to the question (these words were not bolded in participants' conversational prompts page). In addition to the questions themselves increasing the use of these words, the answers also typically involved high rates of test tokens for /æ/. Question (16) typically resulted in at least one story from each participant about the last time they were mad, producing multiple stressed tokens per participant of *last, mad, angry, bad.* Question (17), while not containing any test tokens within the question, often resulted in stories from participants that involved *laughing*, additionally, because participants were high school or college friends, many of these stories took place in *class.* Question (18) straightforwardly produced multiple stressed tokens of *panic* per participant. In this way, each of the four test conditions (MAD, LAUGH, HANG, MANAGE) were heightened by this line of questioning.

In this case, the Topic-Directed questions also had the benefit of being thematically related as emotional-state questions, making these inclusion of these questions a natural as a set. Procedurally, participants were told that I was investigating "language and life in Philadelphia," and that I wanted them to chat for about an hour and a half. Participants were told "you may talk about whatever you like, and here is a list of conversational prompts that you're welcome to use." An hour and a half later, I returned to administer the formal methods, outlined in §4.3.2. Because participants were explicitly told that it did not matter whether they followed the prompts or not, not every participant discussed every conversational prompt. However, each dyad did discuss each of the three targeted emotional-state questions, which were found in pilot interviews to be a very productive set of topics that participants were highly engaged in. Overall, this method produced an average of 238 test tokens per participant from the informal conversation section of the interview (contrast with avg. 170 from the traditional sociolinguistic interviews that comprise the

IHELP corpus). The full conversational prompt list is provided in Figure C.1 in Appendix C.

Formal Methods

After an hour and a half of Topic-Directed Conversation, I returned to the recording room to administer the Formal Methods. These were also designed to target the relevant test items for /æ/. Participants were first given a Semantic Differential task, which was slightly modified to enable responses from both participants. Each participant was given a list of different word pairs. Participants took turns reading off a pair, then discussing what they thought the difference between the two words was, followed by their partner responding with their thoughts. In most cases, this resulted in a light debate over the meanings of each pair, producing multiple stressed tokens of the test items per speaker. In the rare case where the partner simply agreed with the first participant or the participant only gave the meaning of one of the words (e.g. "*mad* is more casual"), I prompted further discussion with pointed questions (e.g., "When would use one vs. the other"). The full list of Semantic Differential pairs is given in Table 4.1.

Mad and Angry	Janitor and Handyman	Strangle and Choke	Stammer and Stutter
Sad and Unhappy	Planet and Asteroid	Valley and Canyon	Damage and Destruction
Glad and Happy	Ran and Jogged	Palace and Mansion	Street and Road
Bang and Crash	Swam and Swum	Pal and Buddy	Pollyanna[8]and Secret Santa

Table 4.1: Semantic Differential prompts.

Following the Semantic Differential, participants were asked to read a word list, also provided in Appendix C. The word list included targeted test /æ/ words, as well as several nonce words from the test conditions to help identify the productivity of participant's /æ/ rules. Participants were asked to read the words down rather than across, and were instructed "some of these words aren't real – just say them however you think they should be said."

[8] *Pollyanna* is a term for Secret Santa prevalent in Irish or Italian Philadelphia neighborhoods

4.4 Analysis of individual speakers

Determining which mechanism of phonological change is driving the allophonic restructuring in Philadelphia relies primarily on our ability to determine whether intermediate cohort speakers' productions align with one or more of the predicted outputs highlighted above.

The main distinction in the output between the three mechanisms is in the distribution of the test tokens. In phonetic incrementation, even in the versions that maintain a distinction between HAND and CAT, all test tokens are expected to be produced from a single distribution located intermediately between HAND and CAT. In spontaneous phonologization, the test tokens as a whole will be drawn from two distributions (both HAND and CAT), but each test word class will itself be drawn from a single distribution (either HAND or CAT), following the underlying system that the speaker is adhering to. In competing grammars, each test word class will be drawn from two distributions (HAND and CAT). The basic questions that we seek to answer with statistical evidence are provided in (19)–(20). Each mechanism of change produces a distinct set of answers to these questions, as shown in Table 4.2.

(19) Are the test tokens, overall, bimodal?

(20) Is each conditioning factor bimodal?

	(19)	(20)
Phonetic Incrementation	no	no
Spontaneous Phonologization	yes	no
Competing Grammars	yes	yes

Table 4.2: Profiles of distributions for test tokens and conditions for each of the three mechanisms of change.

As we will see, the statistical methods for analyzing token distribution are not well set up to answer these questions for unsupervised data. A full analysis of individual speakers will rely on bringing statistical, phonological, and sociolinguistic evidence together to bear on the question of the mechanism of phonological change.

4.4.1 Statistical Evidence

The multidimensional and unsupervised nature of this data means that standard statistical methods are not well set up to address (19) and (20). Most tests of bimodality rely on a label given to each group of tokens, and then asking whether each group has been drawn from the same sample. In our case, we cannot give an *a priori* label to any of the test tokens, since the expected distributions under each mechanism of phonological change range from a probability of 0 to 1 for any label that we may give to an individual observation.

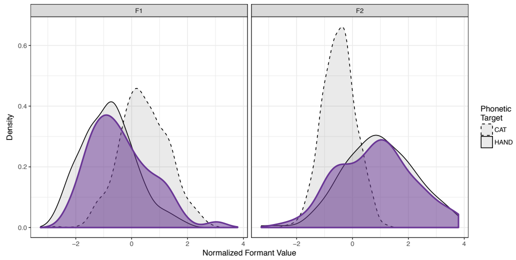
In what follows, I highlight a few statistical methods that in theory would provide some support to our goal of identifying test token observations as being produced by one of the three mechanisms of phonological change. In each case, I test each method on the simulated data (particularly the simulated data represented the expected output of a Cohort 3 speaker) from each mechanism of change outlined above, to test whether each method is able to distinguish data that we know the underlying distribution of.

Multidimensional Kurtosis

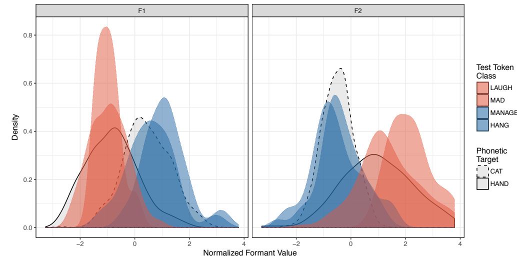
As I have discussed in §4.1.1, unsupervised cluster analysis did not provide a useful tool for distinguishing underlying classification of tokens. Turning to the question of testing modality (unimodal vs. bimodal distribution), I first test the usefulness of a multidimensional kurtosis measure (also known as Mardia's test). Mardia's test in the MVN package in R provides Pearson's adjusted kurtosis values, which are generally interpreted as unimodal (or normal) between the range of -2 and 2, and bimodal outside of this range. Unfortunately, Mardia's test does not provide a reliable distinction between the simulated data for phonetic incrementation (which should produce a unimodal distribution) and competing grammars (which should produce a bimodal distribution), so cannot be used as a reliable measure of bimodality for the data from transitional cohort speakers.

Kernel Density

Kernel density estimation provides an estimate of the probability density function of a variable, essentially providing an output that can be interpreted visually. Using the ks package in R, I fit a

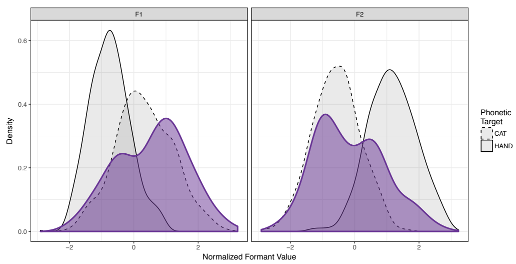


(a) Bimodal distribution of test tokens for PHL speaker.

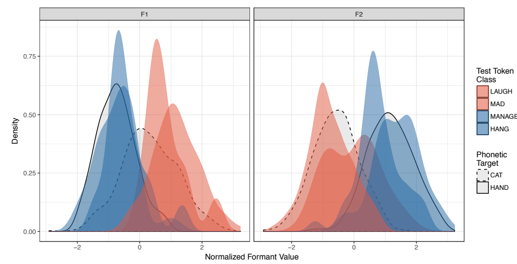


(b) Unimodal distribution of each test conditioning factor for PHL speaker.

Figure 4.13: PHL speaker kernel density plots.



(a) Bimodal distribution of test tokens for NAS speaker.

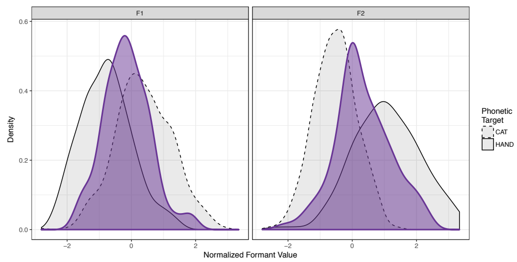


(b) Unimodal distribution of each test conditioning factor for NAS speaker.

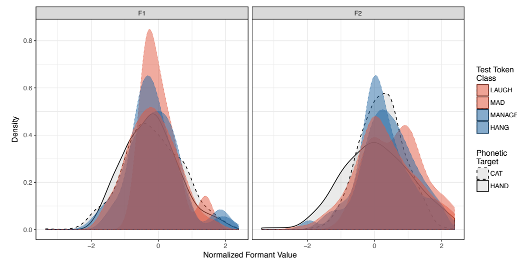
Figure 4.14: NAS speaker kernel density plots.

multidimensional kernel density estimate to each class of test tokens. These estimates can then be compared to each other to test whether the distribution of each class is from the same sample or not. In principle, this would provide a useful tool for testing whether each conditioning factor was equally participating in any changes (as expected for a competing grammars analysis and most phonetic incrementation predictions) or if each conditioning factor was separately affected by the change from PHL to NAS. In practice, the number of observations per conditioning factor for most speakers makes the kernel density plots difficult to lean on as analysis tools. However, the predictions that each mechanism of phonological change make with regards to kernel density are worth briefly discussing. In the final analysis of each individual speaker, I use sociophonological as well as kernel density evidence to classify each speaker as consistent with PHL, NAS, competing grammars, or phonetic incrementation.

Here, a visual plot of the predicted kernel density outcomes for each mechanism of change is



(a) Unimodal distribution of test tokens for phonetic incrementation.



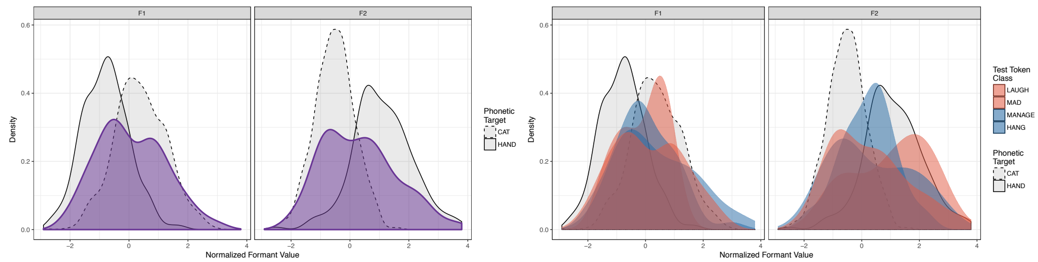
(b) Unimodal distribution of each test conditioning factor for phonetic incrementation.

Figure 4.15: Kernel density plots of predicted output for phonetic incrementation of test tokens.

useful. Figures 4.13 and 4.14 present kernel density plots for the simulated PHL speaker's data and the simulated NAS speaker's data, respectively. Each of these plots comprise a kernel density plot of the test tokens as a whole (a) as well as a kernel density plot of each test conditioning factor (b). The target HAND class and CAT class tokens are also presented in each plot, as a benchmark for each simulated speaker's tense and lax targets. For maximal clarity, each conditioning factor in (b) is colored according to its expected realization under PHL: red for the LAUGH and MAD class, and blue for the MANAGE and HANG class. Figures 4.13a and 4.14a clearly display a bimodal distribution in the test tokens overall. Figures 4.13b and 4.14b display unimodal distributions for each conditioning factor, as either within the tense target or the lax target.

Compare this to Figure 4.15, which presents the kernel density plots for F1 and F2 for the simulated data of phonetic incrementation of the test tokens. Here, I use the variation of phonetic incrementation that results in only the test tokens incrementing, because the stability of the HAND and CAT classes in this variation makes it the most difficult to disambiguate this variation from the outputs of competing grammars and spontaneous phonologization. In Figure 4.15a, the unimodal distribution of all test tokens, produced intermediately between the tense and the lax phonetic targets, is clear. This unimodality also holds for each test condition (Figure 4.15b), in which each test condition also displays a unimodal distribution centered between the tense and the lax targets.

Finally, in Figure 4.16, I present the kernel density results for the predicted output from a Cohort 3 competing grammars speaker. This data comprises 50% PHL-consistent tokens and 50%



(a) Bimodal distribution of test tokens for competing grammars.

(b) Bimodal distribution of each test conditioning factor for competing grammars.

Figure 4.16: Kernel density plots of predicted output for competing grammars speaker.

NAS-consistent tokens. In Figure 4.16a, the bimodality of the test tokens as a whole is clear and reminiscent of the test kernel density plots for the PHL speaker in Figure 4.13a and the NAS speaker in Figure 4.14a. The test conditioning factors, on the other hand, provide a stark contrast to the unimodal distributions for each test condition found in the PHL speaker, the NAS speaker, and the phonetic incrementation speaker. Here, each test condition is bimodal, with 50% of each condition produced as tense and 50% produced as lax.

Kernel density plots for test conditions overall and each test conditioning factor are provided in Appendix D for each transitional cohort speaker analyzed in this chapter, along with that speaker's production plots.

4.4.2 Sociophonological Evidence

Where the statistical methods for determining underlying distribution of test tokens fall short of our goal, the sociophonological evidence provides additional important cues.

Central to a sociophonological account /æ/ is found in the details of speaker production of PHL described Labov (1989). In Labov (1989), PHL-dominant speakers were found to occasionally produce lax forms of traditionally tense PHL tokens, which I've termed *incongruent tokens*. Speakers' rates of inongruent tokens increased during the more formal components of the interview, with the highest rate (15%) found during the reading list. This behavior is in line with the finding of Labov (2001) that PHL speakers downgrade tense PHL tokens but not lax PHL tokens. Taken together, these two findings produce a prediction that speakers occassionally phonetically mitigate up to 15% of their tense tokens, particularly in more formal methods.

From the viewpoint of understanding speakers' systematic production of their language, both the style shifting and the targets of the style shifting provides some important pieces of information: first, that speakers are consistent within their casual style of speech, and second, that they phonetically mitigate a stigmatized tense form to lax in more formal settings, but do not produce traditionally lax forms as tense. From a sociophonological point of view, this provides an important backdrop to the data that follow. We may posit that any speaker producing a clearly tense token of LAUGH or MAD (the tense classes that could only be produced by an underlying PHL system) as well as a clearly tense token of MANAGE or HANG (the tense classes that could only be produced by an underlying NAS system) is exhibiting the operation of two systems within their speech. If such variation is found within the casual portion of the sociolinguistic interview, it can be given more analytic weight, since it is during the most casual speech that speakers behave most systematically. We can add to this the clear prediction drawn from a Competing Grammars framework: not only would we expect tense tokens from PHL and NAS, but we would also expect lax tokens from each system as well.

Here, I take a conservative approach and require a speaker to produce more than 15% incongruent tokens, either produced as lax or tense, in order to be classified as a speaker exhibiting competing grammars. The modified Magnitude Estimation task presented in Chapter 5 finds Philadelphians do not rate tense productions of NAS poorly; therefore, a tense NAS token cannot be used as evidence for an underlying NAS system in the same way that a tense PHL token can be used as evidence for an underlying PHL system. As a result, I simply set a limit of 15% incongruent tokens as a defining limit for a competing grammars speaker. I use this along with the kernel density plots to determine (1) whether the production of test tokens is unimodal and (2) if not, whether the production of test tokens is consistent with a competing grammars hypothesis.

4.5 Results

Here, I will first demonstrate my analyis of two individual speakers before turning to the community-wide pattern. In §4.5.1, I return to the data from Bobby Marx, who is analyzed as an underlyingly PHL speaker who produces fewer than 15% incongruent tokens. In §4.5.2, I analyze the production of "Orange Juice", a 25-year-old transitional cohort speaker from the IMPC data set who produces the highest token count for test tokens of /æ/ and whose production is most compatible with a competing grammars analysis of change. In §4.5.3, I present the results of analyzing every transitional cohort speaker (speakers from the IHELP and IMPC data sets born after 1983) who produce enough data (at least 5 tokens per test condition) to bear on the mechanism of phonological change.

4.5.1 PHL speaker: Bobby Marx

First, I return to the data from Bobby Marx, whose data is ultimately classified as a PHL-conforming speaker. As we've noted above, Bobby produces some incongruent tokens, both as PHL and as NAS. From a social evaluation perspective, the existence of tense tokens from both systems suggests the operation of both PHL and NAS. However, here I take a conservative approach to classifying speakers as competing systems speakers, and use instead the benchmark of whether more than 15% of his test tokens overall are incongruent.

In Figure 4.17, Bobby's test tokens are plotted as text, to enable a sociophonological analysis. His kernel density plots are provided in Figure 4.18a (the kernel density of his test tokens as a whole) and Figure 4.18b (the kernel density of each test conditioning factor). Bobby's kernel density plots for his HAND class (solid line) and CAT class (dotted line) are provided as benchmarks for his tense and lax targets as well. The relative sparsity of Bobby's data results in kernel density plots that are difficult to read, but the main takeaway from these plots is the apparent separation between the red conditions (LAUGH, MAD classes) and the blue conditions (MANAGE, HANG classes)

Figure 4.17 provides more insight into the incongruent tokens in Bobby's production. Several of the lax tokens (*alas, asterisk*) are straightforwardly lexical exceptions to lax for most PHL speakers. Likewise, we see that Bobby produces all tokens of *planet* as exceptionally tense forms (in

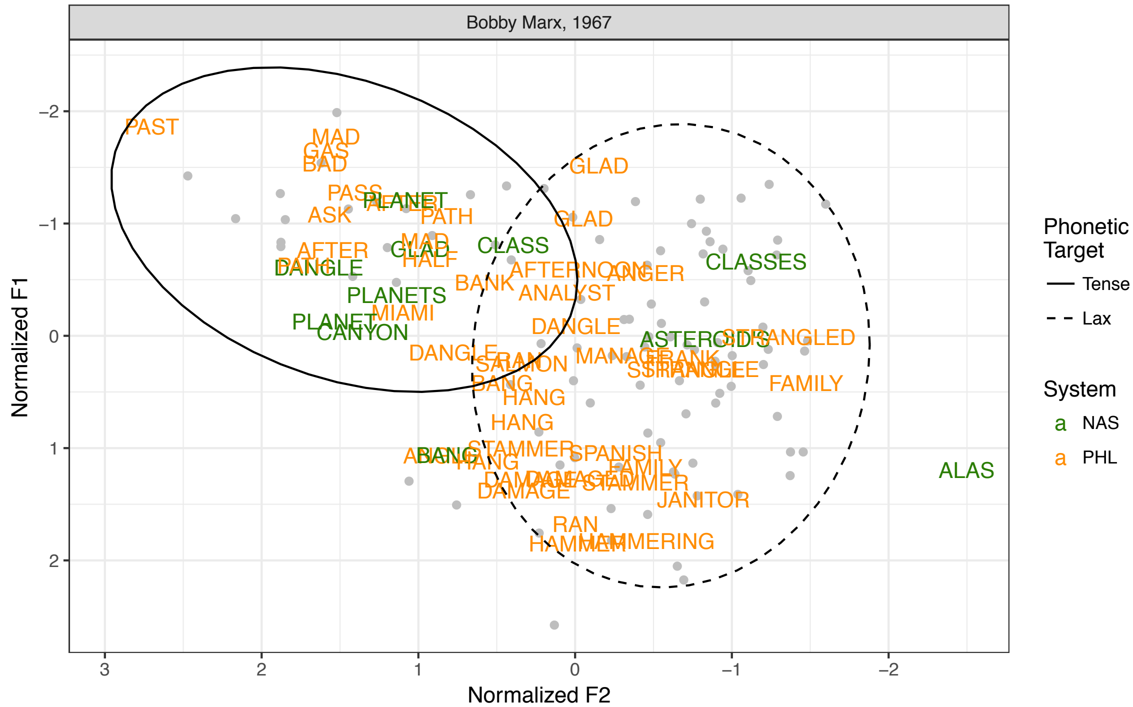
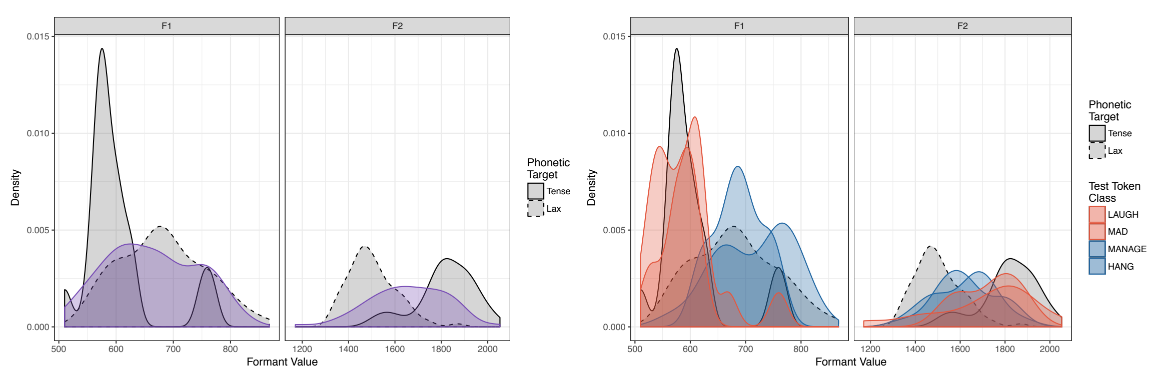


Figure 4.17: Bobby Marx test tokens by word identity.



(a) Bimodal kernel density distribution of test tokens overall.

(b) Kernel density distributions of each conditioning factor.

Figure 4.18: Bobby Marx kernel density plots.

contrast to the rest of his MANAGE class words, which he produces well within his lax target). Bobby represents one of the earliest speakers in our dataset who has adopted *planet* as an exceptionally tense form, in line with Brody (2011)'s findings. Because these tokens are canonical lexical exceptions, I do not count them towards Bobby's overall incongruent count.

Aside from these expected exceptions, Bobby also produces several more interesting tokens consistent with NAS. Both of his tokens of *class* are classified as lax. This appears to be a lexical exception that has been added to his lax list, or is perhaps a single lexical item that has been singled out for phonetic mitigation to lax. This leaves four tokens unaccounted for: *glad, canyon, dangle*, and *bang*. There are three possible explanations for this production. The first is that Bobby has added *dangle* and *canyon* to his exceptional tense list and *glad* and *bang* to his exceptional lax list. The second is that he is exhibiting an early stage of grammar competition between PHL and NAS, with NAS only produced a small percentage of the time. The third is that the classifier misclassified tokens or that these tokens were speech errors. In order to truly disambiguate between these three options, we would need more data from Bobby, including multiple tokens from each lemma.

For the purposes of classifying the data that exists, I count these tokens as incongruent. By token count, this means that Bobby produces six incongruent tokens *dangle, canyon, glad, class, classes, bang* out of 56 total test tokens, meaning that 10.7 % of his tokens as incongruent. As this falls below the threshold of 15%, I classify Bobby as a PHL speaker.

4.5.2 Competing Grammars speaker: Orange Juice

I turn next to the speaker in the IMPC data set with the highest token count, "Orange Juice". Orange Juice is a 25 year old women who graduated from an Open Admissions Catholic high school and the regionally-oriented Drexel University, produces the highest number of /æ/ tokens (894) and the highest number of test /æ/ tokens (331) of all the IMPC speakers.

On the surface, Orange Juice's data (Figure 4.19) is clearly most consistent with a competing grammars profile. Her kernel density plots (shown in Figure 4.20) provide support for this conclusion, particularly in the apparent bimodality in the test token conditions shown in 4.20a. Given what appears on the surface to be clear variation between PHL and NAS in a competing grammars

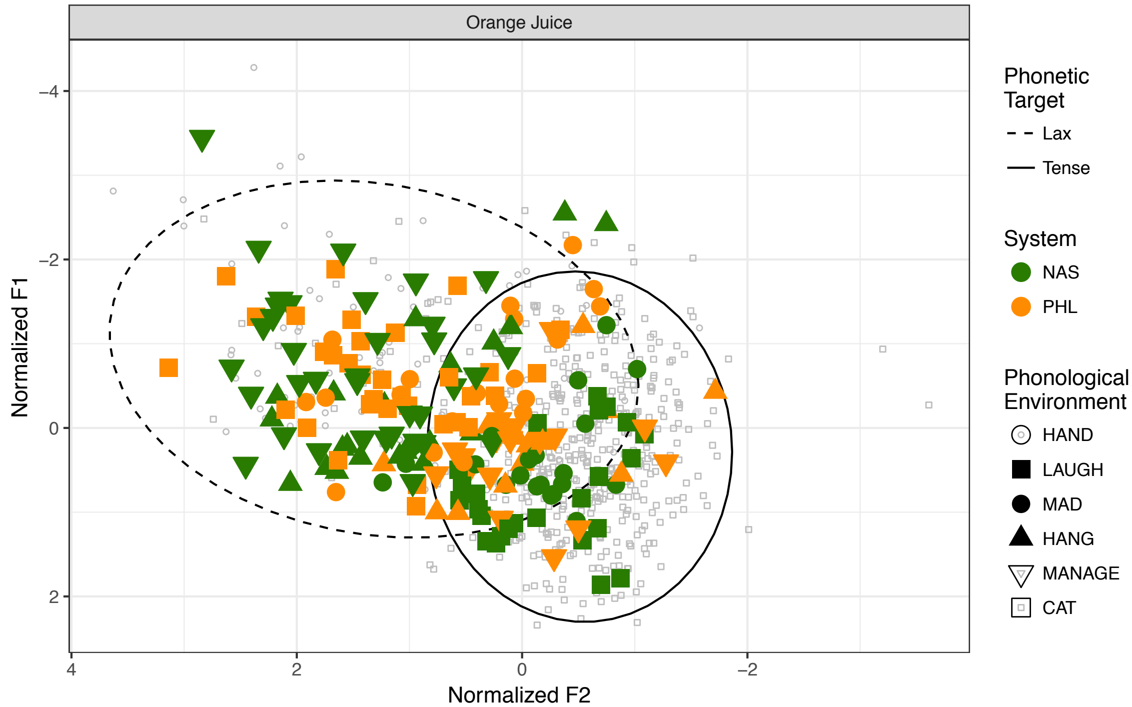
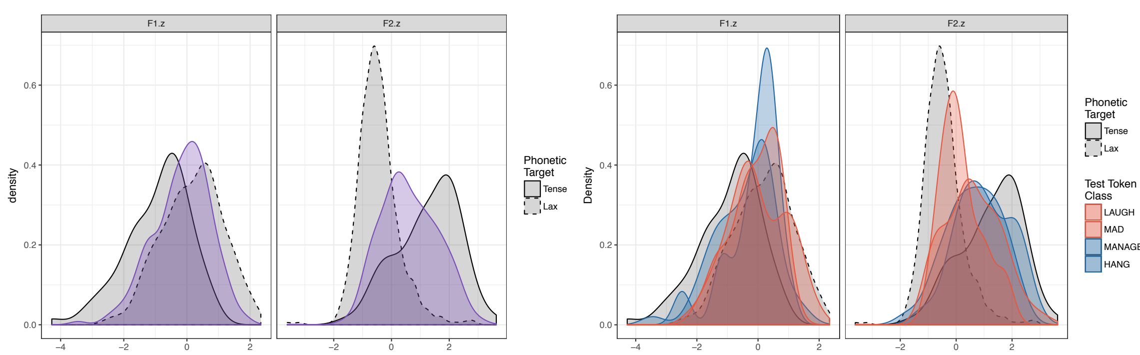


Figure 4.19: Orange Juice production of /æ/.



(a) Kernel density of Orange Juice's test tokens overall.

(b) Kernel density of Orange Juice's test conditioning factors.

Figure 4.20: Orange Juice kernel density plots.

type output, there are two alternative explanations for Orange Juice's distribution that must be ruled out before classifying her as a competing grammars speaker.

Disjunctive rules

Recall that PHL is comprised of a disjoint set of phonological triggers. While I have represented PHL as a single rule with disjoint triggers, it would also be possible to represent the traditional /æ/ system as two separate rules, shown in (21).

(21) a. **PHL$_1$:** $\text{æ} \rightarrow \text{æh} \,/\, \_\left[\ \text{+anterior}\ \right] \cap \left[\ \text{+nasal}\ \right]]\, \sigma$

b. **PHL$_2$:** $\text{æ} \rightarrow \text{æh} \,/\, \_\left[\ \text{+anterior}\ \right] \cap \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix}]\, \sigma$

If speakers represent the traditional input as two distinct rules rather than a single parameter as I have hypothesized, it is possible that the surface level variation found in Figure 4.19 is simply the result of Orange Juice discarding one of the two rules. If, for example, Orange Juice rejected PHL$_2$, she would produce tense /æ/ preceding anterior tautosyllabic nasals (HAND) and lax tokens elsewhere. This means that tokens preceding intervocalic nasals (MANAGE) and velar nasals (HANG) would be produced lax, appearing as surface-level PHL tokens. Orange Juice would also produce lax tokens preceding voiceless fricatives (LAUGH), which would appear as surface-level NAS tokens. This scenario can fairly quickly be ruled out by taking a closer look at Figure 4.19, in which we see instances of each test conditioning factor in both the tense and lax targets, as depicted by token shape.

Lexical diffusion

A final possibility that must be falsified before concluding that we have found competition between PHL and NAS is whether the surface-level variation is simply the result of lexical diffusion. Traditional PHL input requires speakers to memorize a fairy extensive list of lexical exceptions both as tense and as lax. We've also seen evidence that the specific lexical entries are subject to diachronic change, with *planet* joining the exceptionally tense class for many speakers born around 1990 (Brody, 2011) and various words leaving the exceptionally lax class (e.g. *ran, swam, began, and* for speakers born around 1985). This raises the distinct possibility that the variation within conditioning factors found in speakers may actually the result of lexical diffusion into or out of the list of lexical exceptions. For example, if Orange Juice produced PHL but added *janitor* to her list of exceptionally tense tokens, she would produce tense *janitor* and lax *manage*, resulting in what appears on the surface to be variation within the intervocalic nasal conditioning factor. If this she then also added *hang* to their exceptionally tense list and *class* to their exceptionally lax list, she would appear on the surface to produce variation within all conditioning factors that distinguish between PHL and NAS. If, on the other hand, Orange Juice's is the result of competing PHL and NAS, she is expected to produce variation between PHL and NAS within a each lemma.

Competing Grammars in Orange Juice

With her high token count, Orange Juice's data provides the best opportunity to test whether what looks like variation between PHL and NAS is the result of a selective adherence to only one of the PHL constraints, the result of lexical diffusion, or the result of competition between PHL and NAS.

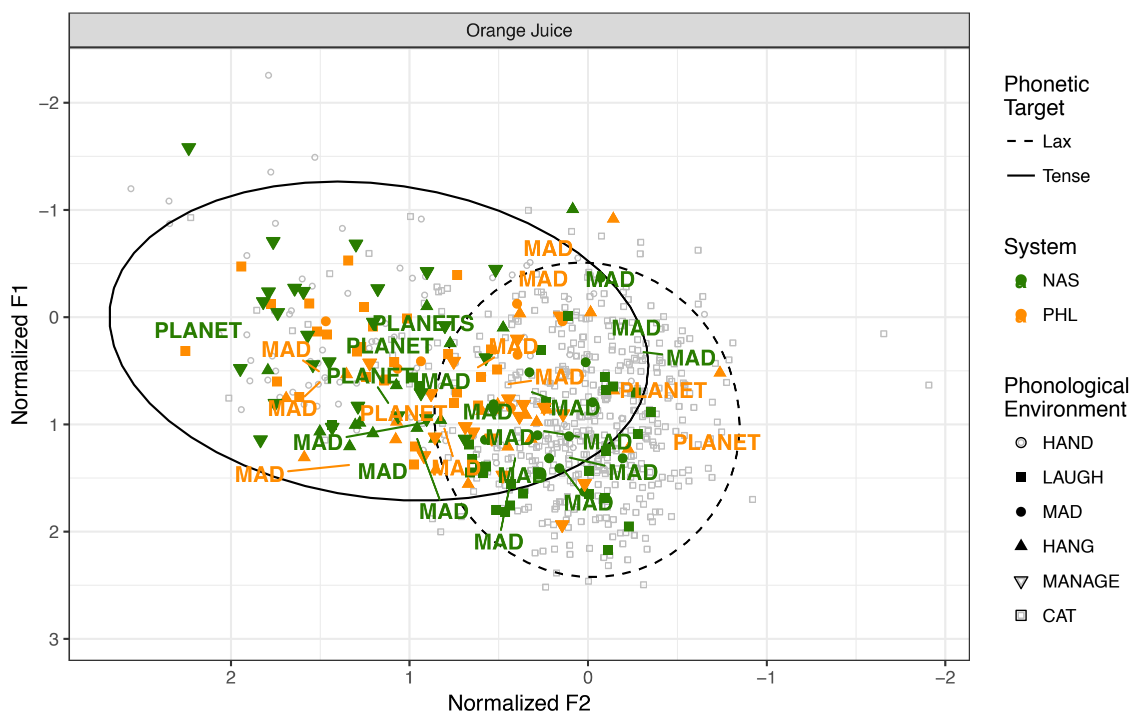


Figure 4.21: Within-lemma variation in the production of Orange Juice.

First, to the disjointed rules. We have already seen both tense and lax forms from each test conditioning factor in Figure 4.19 (see Appendix D for a full presentation of Orange Juice's data). The next possibility to rule out is variation through lexical diffusion. Figure 4.21 presents the wide variation in Orange Juice's production of two lexical items (*mad* and *planet*). Because Orange Juice exhibits variation in all test conditions, and because this variation is not driven by lexical diffusion, her data is best classified as an example of competing PHL and NAS within a single speaker.

Having found a clear example of competing grammars within a single speaker, I now turn to the data from the community as a whole to see whether this mechanism of change is the primary driving force for phonological change across the community.

4.5.3 Analysis of the community

Using the methods described above for Bobby Marx and Orange Juice, I classify each transitional cohort speaker in the IHELP and IMPC data sets who produce at least 5 tokens per test condition according to how their production matches up with the profiles of phonological change outlined above (47 speakers in total). Each speaker's data is also provided in Appendix D, which includes word identity and kernel density plots for each speaker. The majority of transitional cohort speakers conform either to PHL (7 speakers, Figure 4.22) or NAS (19 speakers, Figure 4.23).

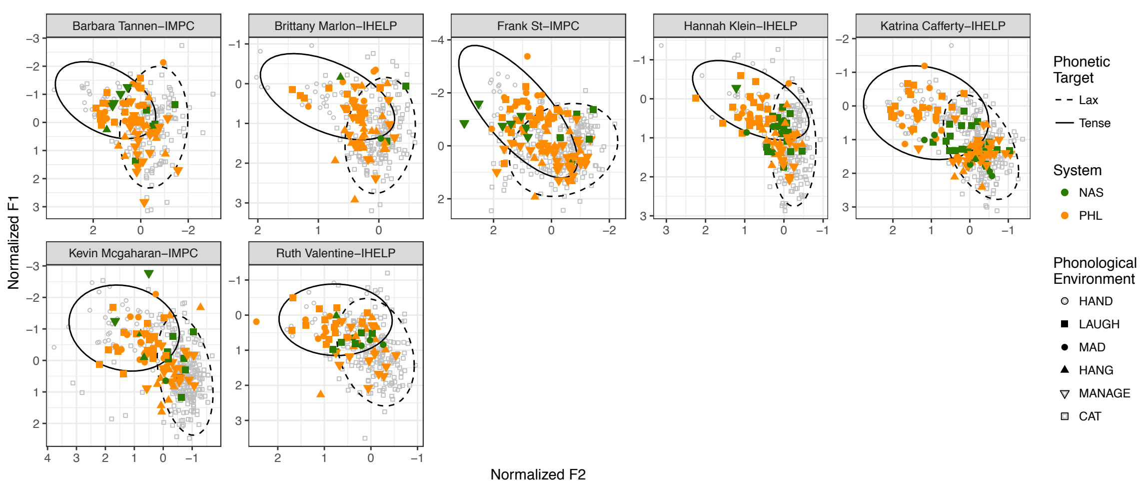


Figure 4.22: PHL speakers.

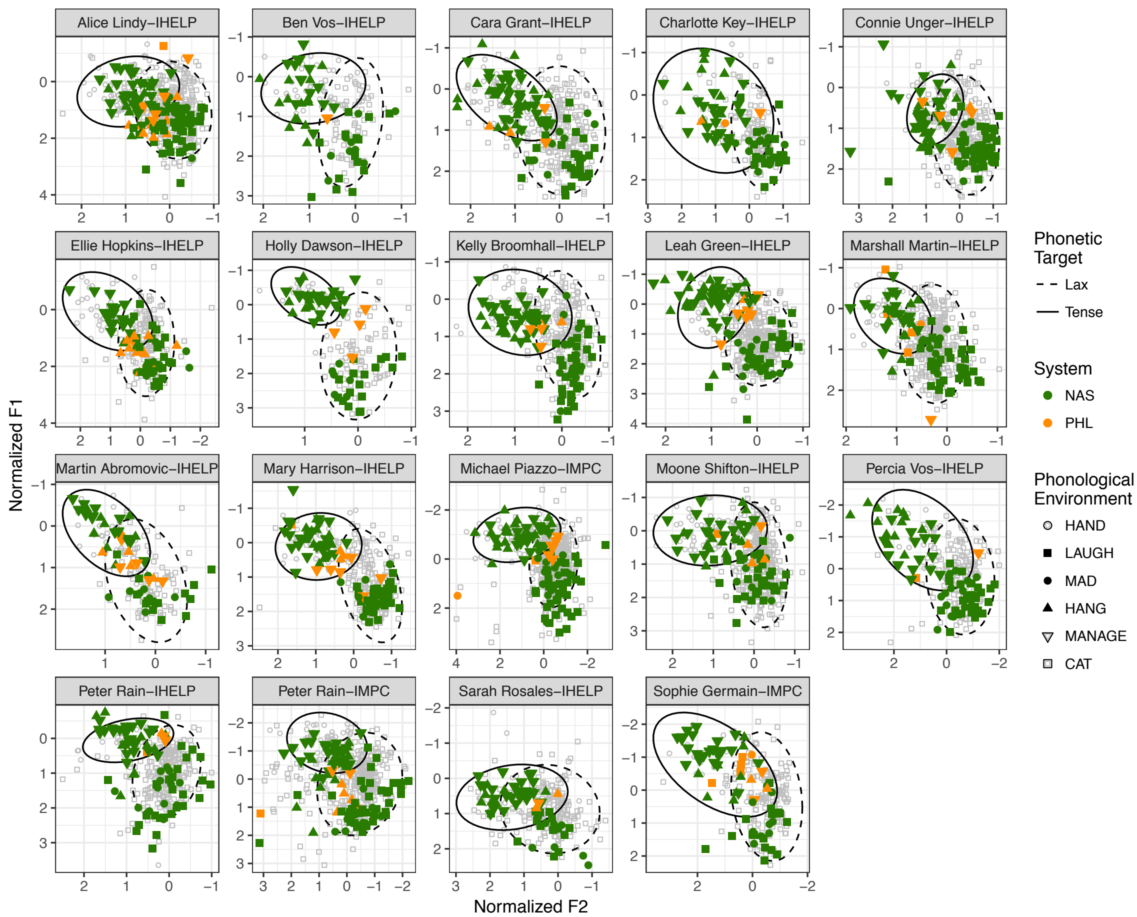


Figure 4.23: NAS speakers.

As discussed briefly in Chapter 1, finding speakers conforming primarily to PHL or NAS does not in itself provide support for any of the proposed mechanisms of phonological change. This data could be consistent with an analysis of some speakers driving phonological change via spontaneous phonologization, but could also simply be the result of the fragmentation of the speech community. A closer look at the educational history of each of these speakers reveals that their production is most likely the result of their school either not yet having undergone the change to NAS or having already completed the change. With the exception of Kevin, all of the speakers in Figure 4.22 attended an Open Admissions Catholic school, which as we have seen in Chapter 2 are conservative strongholds for PHL. Kevin emerges as somewhat of an outlier: he produces a PHL system despite having attended a Special Admissions non-Catholic high school. I note, however, that the neighborhood he grew up in (in South Philadelphia) is a stronghold for traditional PHL, and that he attended an Open Admissions Catholic middle school, providing an avenue for his acquisition and maintenance of PHL.

Similarly, most of the speakers classified as NAS speakers in Figure 4.23 were graduates of Special Admissions non-Catholic schools, with six exceptions. Five of these exceptions (Alice Lindy, Michael Piazzo, Moone Shifton, Peter Rain, and Sophie Germain) attended Special Admissions Catholic schools, which were identified in Chapter 2 as a segment of the schooling system that produces both NAS and PHL speakers. Here, these five speakers represent graduates of Special Admissions Catholic schools who have adopted the new NAS system. The sixth exception, Marshall Martin, attended an Open Admissions suburban public school, which we have previously identified in Chapter 2 as socially similar to the Special Admissions non-Catholic schools within the city. The educational histories of speakers classified as PHL or NAS in Figures 4.22 and 4.23 reveal that these speakers are not driving phonological change, but rather represent a segment of the population that either has not yet undergone the change from PHL to NAS or has already completed this change.

There are an additional fourteen speakers (Figure 4.24) who produce outputs most consistent with a competing grammars analysis of change. Seven of these speakers (Jacob Ambrose, Elizabeth Rina, Steve Rina, Ariana Tocci, Orange Juice, Speedy Racer, Wendy Juice) graduated from

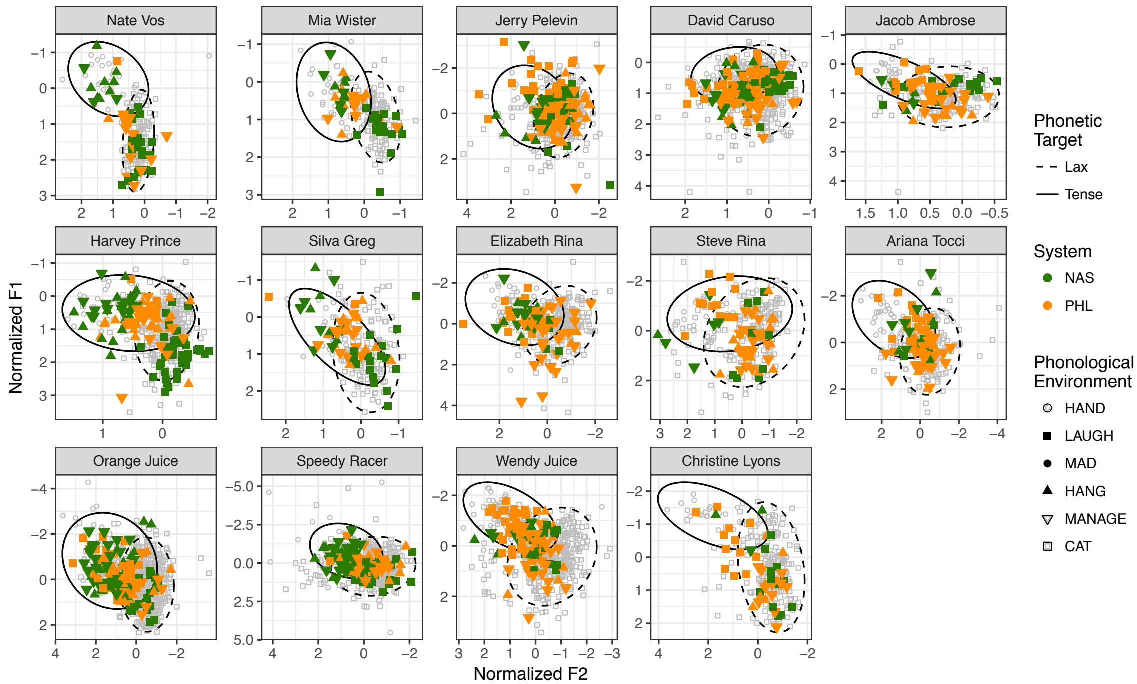


Figure 4.24: Competing grammars speakers.

Special Admissions Catholic schools, which is that section of the school system that has been already identified as the most vigorously changing in Philadelphia. Several of these speakers, Nate Vos and Christine Lyons, has already been discussed in some detail in Chapter 2. While Nate appears exceptions, as he attended Masterman middle school at the time of his interview, he had only been in that environment for a few months and had received mixed-system input from his father (Harry). Nate's production of a competing grammars output suggests that it is the result of his not yet settling on a NAS output. Christine is exceptional for the opposite reason: she is the graduate of an Open Admissions Catholic high school who nevertheless produces a competing grammars output. Recall, however, that Christine also attended Penn – a nationally oriented Ivy-league university – which is likely to have had an impact on her production. There are a few other speakers whose productions are exceptionally conservative or innovative given their educational background. Mia Wister, David Caruso, Jerry Pelevan, Silva Greg, and Harvey Prince were graduates of Special Admissions non-Catholic high schools. Jerry and Harvey, however, also attended Open Admissions Catholic middle schools. Likewise, Jerry and Silva attended the locally-oriented Temple University for college. The influence of traditional middle schools and universities for these speakers may provide some insight into their ability to retain a competing grammars system. Mia Wister and David Caruso emerge as surprising conservative exceptions. Both Mia and David attended Special Admissions non-Catholic middle and high schools, before attending Penn for college.

In the detailed examination of speakers' educational histories provided above, we have found that Open Admissions Catholic schools largely produce PHL-conforming speakers while Special Admissions non-Catholic schools largely produce NAS-conforming speakers. Special Admissions Catholic schools are the main schools that produce competing grammars speakers. This is unsurprising, given the results in Chapter 2: this is the segment of the broader speech community that bridges the Special Admissions non-Catholic schools and the Open Admissions Catholic schools and is most likely to have contact with enough NAS and PHL speakers to adopt both systems (I argue in Chapter 6 that a competing grammars speaker requires between roughly 46% and 54% input from both systems during acquisition in order to acquire both as plausible productive rules). We see a few speakers (Mia, David) who trail their peers in the adoption of NAS, as well as one speaker (Christine) who is on the vanguard of this change given her educational history. Overall, the educational histories and production of speakers in Figures 4.22 through 4.24 provide a compelling argument that the change from PHL to NAS is occurring via the mechanism of competing grammars.

4.5.4 Unclassified Speakers

There are a few speakers, all found in the IHELP data set, whose data is not easily classified by the combination of statistical and sociophonological methods that I employ here. Six of these speakers, shown in Figure 4.25, display overall conformity to either NAS or PHL in the non-overlapping parts of their tense and lax targets. For these speakers, the proportion of incongruent tokens exceeds the 15% threshold while they also produce bimodal distributions of their test tokens. Under the classification system I outline above, these speakers would be classified as competing grammars speakers. However, due to the distribution of their tokens, I remain skeptical of analyzing them as competing grammars speakers.

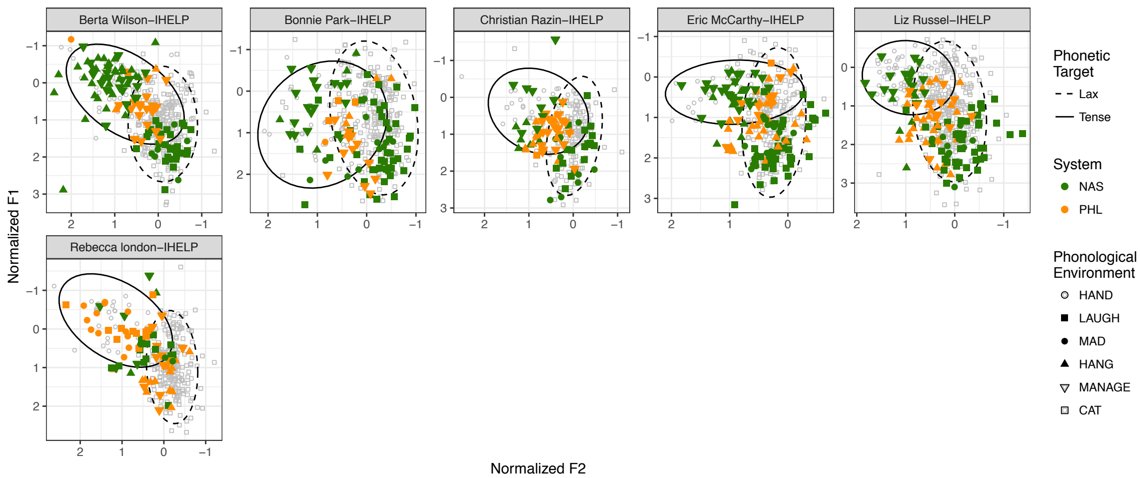


Figure 4.25: Largely PHL or NAS speakers who nevertheless produce more than 15% incongruent tokens.

In addition to the six speakers analyzed above, we find one speakers whose production is consistent with the profile of phonetic incrementation of the test tokens: Jake Stone. Jake's production was already discussed as a community outlier in Chapter 2. As previously discussed, Jake's sociolinguistic background suggests that he is not a transitional cohort speaker producing phonetic incrementation, but rather is simply phonetically mitigating the stigmatized tense production of his underlying PHL system. Jake attended Masterman high school at a time when his peers already demonstrate conformity to NAS, meaning that his phonetically mitigated production of tense PHL tokens is not a likely driver of this phonological change. As in Chapter 2, I analyze Jake as having an underlying PHL system, but phonetically mitigating his traditionally tense tokens to lax, in a sociolinguistic avoidance of a stigmatized form.

Finally, I turn to the last unclassified speaker, Carlos Santana. Carlos' data largely conforms to the NAS system, though he has a few PHL-congruent tokens. He shows a few non-extreme tokens of traditional tense PHL forms, *path, gas, mad, after.* The rest of his PHL-consistent tokens are lax productions of MANAGE and HANG class words. Carlos' production is not consistent with any of

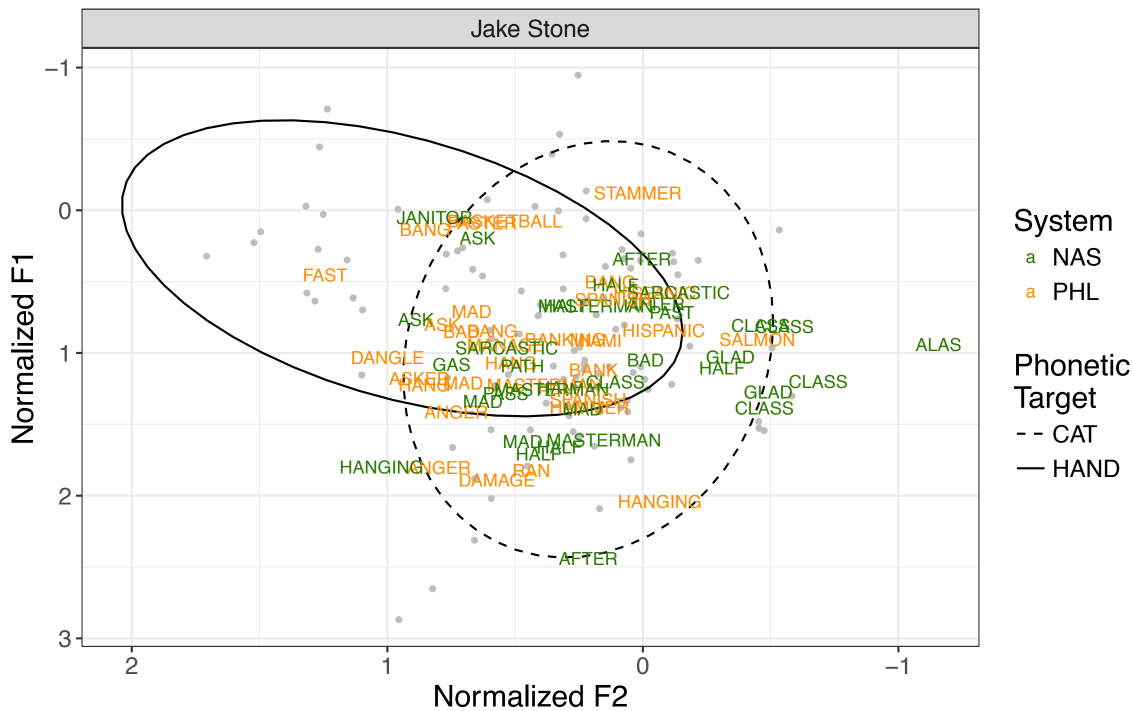


Figure 4.26: Phonetic mitigation in Jake S.

the profiles of phonological change outlined above. He does not adhere to NAS or PHL entirely, as a mechanism of spontaneous phonologization predicts. He also does not produce clearly tense and lax tokens of each conditioning factor, as a mechanism of competing grammars predicts. Carlos' production also does not fit a model of phonetic incrementation, with tokens produced clearly within and even more extreme than his underlying phonetic tense and lax targets.

So what can be said about Carlos and his unexpected production? One potential answer to Carlos' data is that he produces a combination of grammar competition and phonetic mitigation. As I will argue in Chapter 5, younger speakers in Philadelphia acquire two evaluation systems for /æ/ tokens. Any token consistent with NAS is rated highly, as is any lax token of PHL. PHL-specific tense tokens, on the other hand, are rated poorly by younger speakers. If Carlos underlying produced competition between PHL and NAS, then added a filter of phonetically mitigating his PHL-tense tokens (LAUGH and MAD class tokens), the expected output would be as in Figure 4.27.

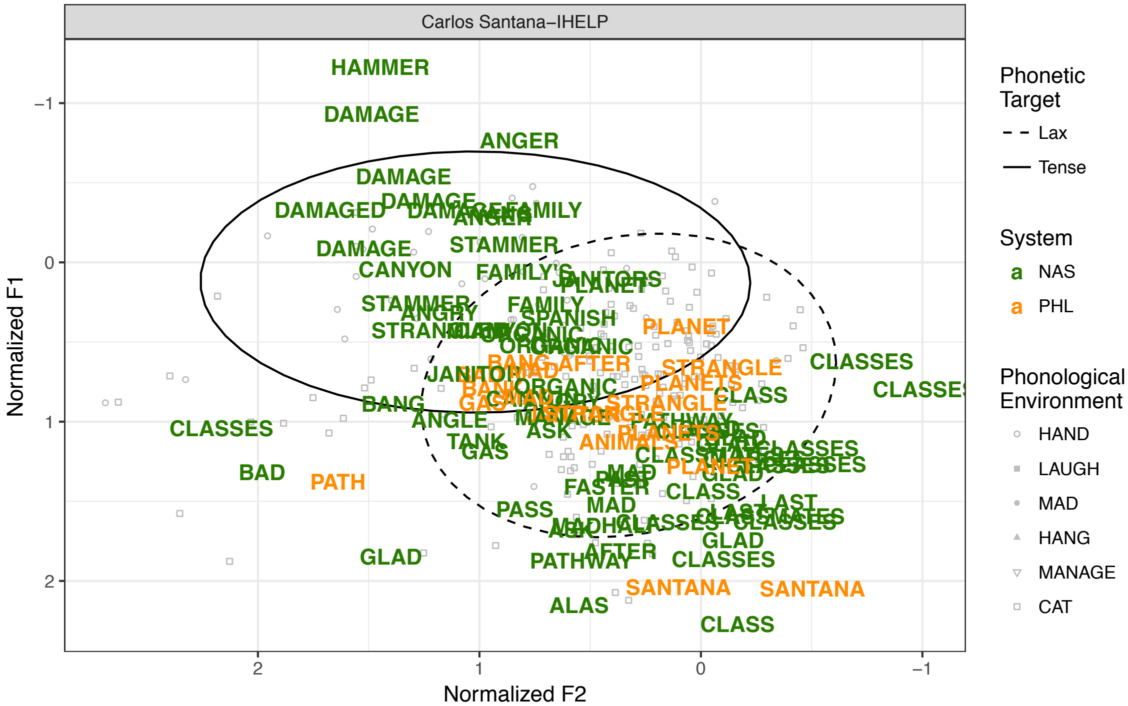


Figure 4.27: Phonetic mitigation of tense PHL tokens in a mixed-system speaker (Carlos Santana).

4.6 Discussion

Throughout this chapter, I've appealed to a number of pieces of evidence to identify what the individual speakers in the IMPC and IHELP data sets are doing. I find that most speakers adhere to either PHL or NAS, but that a sizeable number of speakers also produce variation between the two systems. By closely analyzing the variable speakers with the highest token count, we find that all test conditioning factors exhibit both PHL and NAS productions, and furthermore that variation is even found within the lexical exception lists. This suggests two important outcomes.

First, that grammar competition between PHL and NAS is the mechanism by which this change is occurring, at least in the transitional cohort speakers analyzed here. This is competition between two outcomes of a single "allophonic system" parameter that selects between PHL and NAS each time the speaker goes to produce a token of /æ/. This parameter governs a number of conditioning factors, which all exhibit variability. If our data included longitudinal data from transitional cohort speakers, we would expect these distinct parameters to exhibit a constant rate of change, in accordance with the Constant Rate Hypothesis (Kroch, 1989), also termed the Unity Principle when used to refer to phonological constant rate (Fruehwald, 2013).

Second, we find that the lexical exceptions exhibit variability at roughly the same rates as the phonological conditioning factors. This suggests that the sets of lexical exceptions are cognitively stored under the same single parameter rather than externally to that parameter. It appears that speakers are not producing lexical exceptions and then applying a variable rule which selects whether the non-exceptional forms are produced under PHL or NAS. Instead, we find that the variable rule selects each system wholesale including its lexical exceptions. This finding has important consequences for the structure of the grammar, meaning that any lexical specificity to a phonological process is stored as a component of the phonology itself. This is, in fact, in line with the predictions of the Tolerance Principle (Yang, 2016). The concept behind the Tolerance Principle is that it is a calculation of whether it is more efficient to memorize all lexical items or some lexical items as well as a rule. This is based on the premise that under a rule scenario, all lexical exceptions are processed *before* the rule, in order of lexical frequency. In the strongest formulation of this theory, this is proposed to be an actual model of word production, whereby speakers run through first the list of lexical exceptions then the productive rule any time they go to produce a word.

The data presented in this chapter provide surprising support for this model. If it is the case that lexical exceptions must be serially processed before a regular rule, it follows that that speaker's cognitive representation of that rule includes the lexical exceptions as well. This means that, phonologically, however we represent productive processes must include the possibility of storing lexical exceptions as part of that process as well. Here, I have focused on the most frequent lexical exceptions *mad, bad, glad*, which have also been amplified by the interview methods. As outlined in Chapter 3, I interpret all lexical exceptions to the productive rule in PHL to be stored as a series of exceptions. Given the results reported in this chapter, I would expect *all* instances of lexical exceptions to the productive PHL rule for a given speaker to exhibit the same proportions of variability as do *mad, bad, glad*, since these additional exceptions would also be stored in a similar list preceding the productive rule.

(22) **PHL:**

1. IF w = *and* THEN /æ/ → lax
2. IF w = *can* THEN /æ/ → lax

   ...

39. IF w = *gaffe* THEN /æ/ → lax
40. $\text{æ} \rightarrow \text{æh} \;/\; \_\begin{bmatrix}\text{+anterior}\end{bmatrix} \cap (\begin{bmatrix}\text{+nasal}\end{bmatrix} \cup \begin{bmatrix}\text{-voice} \\ \text{+fricative}\end{bmatrix})]\,\sigma$

(23) **NAS:**

1. $\text{æ} \rightarrow \text{æh} \;/\; \_\begin{bmatrix}\text{+nasal}\end{bmatrix}$

The addition of lexical exceptions to the rule itself requires a slight modification of the notation for PHL, which is argued for in Chapter 3 and reproduced in (22). Here, PHL is comprised first of the lexical exceptions, followed by the regular productive rule. NAS, then, as any other productive phonological rule, also carries the potential for its own list of lexical exceptions, as shown in (23). For the majority of NAS speakers, these potential lexical exceptions list will remain empty, though we do find some speakers in the IHELP corpus whose primary production is NAS but who retain a lexical exception (e.g., in the speaker's own name *Hannah*). In such a case, this speaker's formulation of NAS would be as in (24).

(24) **NAS:**

1. IF w = *Hannah* THEN /æ/ → lax
2. $\text{æ} \rightarrow \text{æh} \;/\; \_\begin{bmatrix}\text{+nasal}\end{bmatrix}$

In other words, our *Hannah* lexical exception speaker is not producing variation between NAS and PHL (using PHL only and every time she says *Hannah*), but rather producing a single system NAS that has a lexical exception. When PHL and NAS are in competition within a single speaker, as found in the competing grammars speakers in Figure 4.24, it is the entirety of (22) and (23) that is in variation. This means that a competing grammars speaker produces tense *mad* ([meːᵊd]) due to the lexical exceptions listed in (22) when PHL is selected, but lax *mad* ([mæd]) when NAS is selected, accounting for the variation found in Orange Juice's production of *mad*.

Here, we have found data suggesting that not only is the change from PHL to NAS in Philadelphia driven by the mechanism of competing grammars, but also that what we consider to be an allophonic rule or an allophonic system is best represented as a single unit containing the possibility for lexical exceptions, as in (22). Finally, it is worth reiterating the major point drawn on in this chapter that an identification of individual speaker's production is not fully possible without an understanding of the sociolinguistic facts of the speech community as a whole.