

# Representational Momentum May Explain Aspects of Vowel Shifts

Samarth Swarup<sup>1</sup> and Corrine McCarthy<sup>2</sup>

<sup>1</sup>Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061

<sup>2</sup>Linguistics Program / Department of English, George Mason University, Fairfax, VA 22030  
swarup@vbi.vt.edu, cmccart6@gmu.edu

## Abstract

We present a computational model of vowel shifts, applied in particular to the Northern Cities Vowel Shift. Our model incorporates several empirically-derived rules of vowel change. The key aspect of this model is the use of representational momentum, which, we argue, explains multiple observed features of the shift. We compare our model with data on the Northern Cities Shift spanning more than a century and show that, when representational momentum is included, the results of the model match the data well.

## Introduction

Language is a fundamentally social complex system, exhibiting organization at multiple spatial, temporal, and structural levels (Beckner et al., 2009).

Language changes at each of these levels, at various time-scales. For example, syntactic changes tend to be very slow, with significant changes taking centuries, while lexical changes can be very fast, often sweeping across social groups in months or years. This is perhaps intuitive, because a change to the syntax signals a change to the entire language, such as the transition from old to modern English.

Some changes, however, take place on an intermediate timescale of ~100-150 years. This is particularly true of phonological changes, such as vowel shifts.

From the point of view of collective dynamics, vowel shifts are interesting for multiple reasons. They are thought to be initiated when two populations come into contact that have vowel systems shifted with respect to each other. This sets off a complex series of changes as people (mostly sub-consciously) adjust their vowel systems in social interactions. Thus, a vowel shift can be seen as convergence on a shared vowel system through a social contagion process.

However, the time-scale of the shift, being over a century, is longer than the lifespan of most humans; this entails that multiple generations of speakers participate in this sound change. Additionally, for the most part, people only adjust their vowel system at a young age, typically until reaching adulthood. This raises the question, what keeps the vowel shift going from one generation to the next, and why do

these shifts appear to stop once they have reached a certain point?

In the present work, we develop an agent-based model of the mechanism of vowel change based on empirical observations by sociolinguists over many years. These observations are incorporated into our model as a system of constraints and update rules that try to preserve these constraints. One key addition we make to the model is to include representational momentum, which is a psychologically documented phenomenon wherein people's memories of sequences of tones ordered by pitch exhibit overshoot. We demonstrate that if we include momentum in the mechanism of individual vowel change, it can account for the long-term population-wide shift in vowels.

The mechanism of vowel change in our model is a phenomenon known as *accommodation*, which is an attempt on the part of a hearer to adjust his vowel system to match the speaker's.

We apply our model to data on the Northern Cities Vowel Shift (NCVS), and show that when representational momentum is included, the results of the model match the data well, while they do not when representational momentum is not included.

The rest of this article is organized as follows. We first present a brief overview of the Northern Cities Shift and the data. Then we discuss prior efforts to model vowel shifts. After that we present our model and experiments with and without representational momentum. We do a statistical analysis of our results to show how they compare with the data, and we end with a discussion of aspects of the shift that are and aren't explained by our model.

## The Northern Cities Vowel Shift

The NCVS is a sound change that is currently affecting the vowels of speakers in the cities along the Great Lakes region of the United States (Labov et al., 2006, henceforth LAB). Its evolution over the last 100 years in Chicago, the largest of the Northern Cities, has recently been documented by McCarthy (2009, 2011).

The NCVS is an example of a chain shift: a series of in-

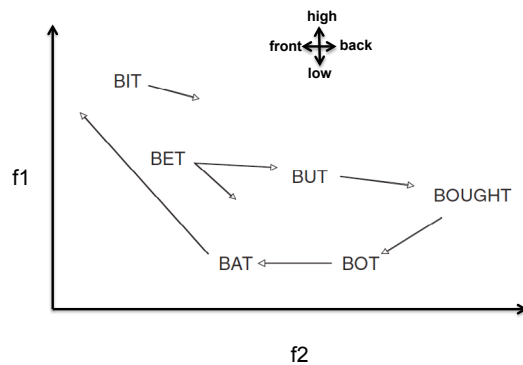


Figure 1: A schematic of the Northern Cities Vowel Shift.

tegrated movements of two or more phonemes. In a chain shift, sounds change in their place of articulation in order to avoid merger and maintain contrast (Labov, 1994; Martinet, 1955).

Consider the example in eq. (1), a minimal chain shift involving two elements that shift in response to one another:

$$/A/ \rightarrow /B/ \rightarrow \quad (1)$$

If the entering element, /A/, shifts first, and /B/ shifts to avoid merger, it is called a push chain. If the exiting element, /B/, shifts first, and /A/ shifts to occupy vacant space, it is called a pull chain (Labov, 1994, 118-119).

The shifting elements of the NCVS are illustrated in figure 1. The axes plot the first two formants (peaks in the frequency spectrum) of the vowels, with the first formant (f1, referred to as “vowel height”) being on the y-axis and the second formant (f2, referred to as “vowel backness”) on the x-axis. Plotted this way, the diagram corresponds roughly to the position in the mouth where each vowel is articulated. Using the first two formants to describe the vowels is standard practice in phonetics (Labov, 1994). Each word corresponds to a vowel class, e.g. BAT represents the vowels of *bad*, *trap*, *cash*, *happy*, etc. Note that front is to the left and back is to the right. The stages of the NCVS, as proposed by LAB and Labov (1994), are as follows:

1. BAT was the first to shift. It went from its traditional, low-front position (IPA [æ]) to one where the nucleus is raised to mid-high position, and followed by an inglide. Pronunciation varies among [eə] ~ [ɪə] ~ [iə].
2. BOT moves forward from the low-back position ([ɑ]) toward the position vacated by low-front BAT.
3. BOUGHT lowers and moves to the front to occupy the space vacated by BOT.
4. BET moves backward toward [ʌ], which is occupied by BUT. BET may also lower toward low-central position.
5. BUT moves backward to avoid collision with BET.
6. BIT lowers to fill the space vacated by BET.

## Chronology of the NCVS

Not all authors concur with Labov’s (1994) and LAB’s ordering of events in the NCVS. Citing data from recordings of speakers born in the 1890s, McCarthy (2009) argues that fronted BOT (“front” as in fig. 1) is robust, but raised BAT is absent from the oldest speakers (see also Thomas, 2001). Evidence for BAT raising first appears in those speakers born in the 1910s. BAT shows rapid shifting from a low position to high position in a span of only about 20 years. McCarthy’s and Thomas’s chronology would mean that the NCVS started as a push chain, with BOT pushing BAT. We will address this issue further in our experiments.

## The apparent-time construct

LAB’s chronology is based on interviews with speakers of different ages. LAB rely on the apparent-time construct: the assumption that language stabilizes after early adulthood, which means that by sampling speakers of different ages, it is possible to observe language change. Sampling the speech of a 60-year-old is the same as sampling a 20-year-old 40 years ago. Sampling the speech of a 40-year-old is the same as sampling a 20-year-old 20 years ago.

In contrast, McCarthy (2009) and Thomas (2001) incorporate real-time data into their analyses, by analyzing recordings made in the 1960s-1980s of speakers who were born prior to 1900.

## Linguistic data for the present study

The present study incorporates both real- and apparent-time linguistic data in order to document a span of 100 years of sound change in Chicago. All speakers were born and raised in the Chicago area, and had parents also from the area.

The real-time data come from archived recordings interviews with six Chicagoans born between 1890 and 1919. Four were gathered as part of the Dictionary of American Regional English project, and two were drawn from Studs Terkel’s interviews, which were digitized by the Chicago History Museum.

The apparent-time data come from sociolinguistic interviews with 35 Chicago-area residents. These speakers read from a list of words containing all the elements of the NCVS.

Speakers were divided into 5 groups based on their birth year. Number of speakers for each age group is as follows: 1890-1910: 3, 1911-1930: 4, 1931-1950: 8, 1951-1970: 10, and 1971-1990: 16.

Vowels were analyzed for f1 and f2 at the point of inflection or midpoint of the steady state, as described in LAB. f1/f2 measurements were normalized using Labov’s G method (Thomas and Kendall, 2007) in order to control for differences in vocal tract size between men and women as much as possible. The data are plotted in fig. 2.

Despite the sample sizes being quite small for the oldest groups, we can make some general observations from the data. First, though Labov claims that BAT raised before

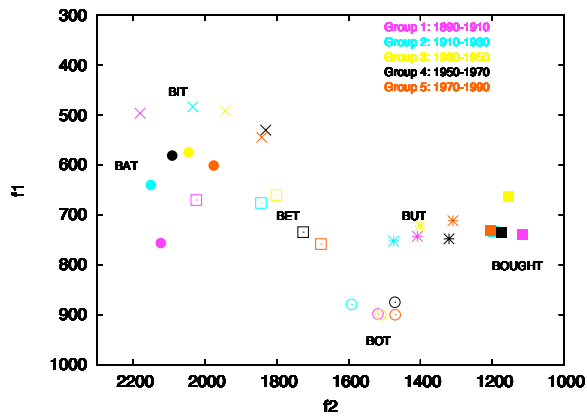


Figure 2: Mean positions of the vowels at various stages of the shift. Note that the directions of the axes are reversed. This is standard practice in the sociolinguistics community. In our model, we will use an idealized range of 0-100 for each axis and they will extend in the usual directions.

BOT fronted, we see that BAT continues to shift after BOT has apparently stabilized. We attempt to address this controversy with our model. Second, we see that BOUGHT doesn't shift much at all, contrary to the description of Labov for the Northern Cities region as a whole. Third, BET and BUT show significant movement, and BUT and BOUGHT get quite close to one another. Fourth, though not previously suggested by McCarthy, there could be an interaction between the movements of BAT and BIT. The mechanism of this last movement will remain an open question. It is as yet unaccounted for by our model. We will address all the other aspects, as well as the more general question of the time-scale of the shift, through our model.

### Related Work

While the vowel system is computationally well-studied (e.g., Joanisse and Seidenberg, 1997; de Boer, 2001; Dras and Harrison, 2003), there has been relatively little work on modeling vowel *shifts*. What work there has been has focused on different aspects of the phenomenon.

Ettlinger (2007) modeled chain shifts along a single axis (the primary frequency, or  $f_1$ ) using an exemplar-based model with just two vowels. In his model, the vowel space is divided into Voronoi cells so that any perceived utterance is categorized as the perceptual prototype in the hearer's vowel system to which it is closest. His primary insight is that if the position of a prototype vowel changes for some reason, it will change the boundaries of its Voronoi cell and cause a consequent change in the position of the prototype in adjacent cells (since the prototype is always the centroid of the cell). This is effectively a chain shift. Note that this model does not directly depend on notions like accommodation.

The main problem with this model is that it does not work

in two dimensions. If we include both  $f_1$  and  $f_2$ , then the alterations in the Voronoi diagram can be much more complex when one of the vowels shifts, and it is hard to reproduce exactly the shift that is seen empirically.

Stanford and Kenny (2012) use a similar model, with three vowels, but their focus is different. They address the question of whether there are differences between child vowel acquisition (by *transmission* from parents and others) and adult vowel change (by *diffusion* through interactions between adults). Their model does not examine the mechanisms by which a chain shift happens. They essentially enforce a vowel shift to study the differences due to frequency of contact between agents with different vowel systems.

Lakkaraju et al. (2012) have the model most similar to ours, though they also restrict attention to change along a single axis in a discrete setting. They also use accommodation to account for change, with two constraints: a phonetic differentiation constraint and a total ordering constraint (which says that there is a total ordering on the vowels, preventing them from leapfrogging each other). The present work extends that model in several ways: by considering the full two-dimensional vowel space, in a continuous setting, with arguably more realistic constraints, and a closer comparison with data.

## The Model

### Convergence in vowel systems

The introduction of a new, incoming linguistic variant such as a shifted vowel creates an opportunity for variation which may ultimately lead to sustained language change as in the NCVS. Linguistic innovation spreads via face-to-face interactions within social networks, and the degree of language change is mediated by a variety of factors, as described below.

According to Communication Accommodation Theory (CAT; Giles and Coupland, 1991), the motivation for convergence for or divergence from their interlocutors is the desire to achieve an optimal degree of social distance. Trudgill (1986), however, proposes that those vowels that are undergoing change in progress might be more salient, and therefore more likely to show accommodation. Goldinger (1998) proposes that convergence may be an automatic cognitive reflex that results from past experience and the type of information that has been stored as exemplars. Under the latter view, accommodation does not depend on social factors.

Babel's (2009) study of accommodation and the California Vowel Shift suggests that different vowels exhibit different degrees of convergence. In an experimental setting, the low vowels (i.e., BAT and BOT/BOUGHT) showed more accommodation in  $f_1/f_2$  values than non-low vowels (i.e., BIT, BOAT, BOOT), perhaps because of the large range of phonetic realizations between stressed and unstressed vowels that result from having to raise and lower the jaw. Social factors (i.e. attitudes and affinity for the ethnic group of the

interlocutor) were shown to mediate the degree of convergence in these experiments.

In Chicago, Herndobler (1977) argues that the spread of raised BAT during the mid-20th Century may have resulted from the perception of urban sophistication associated with the raised variant. This claim would be consistent with socially-motivated accounts of accommodation. More recently, McCarthy (2011) shows that some speakers, especially college-educated ones, have negative stereotypes associated with raised BAT. As social perception of shifted vowels goes from favored to stigmatized, one might expect vowel shifts to peak and retreat. None of the other Chicago vowels implicated in the NCVS appear to be consciously associated with either prestige or stigma.

To summarize, various social and asocial explanations for why speakers converge toward new vowel targets have been offered, and all vowels may not undergo the same amount of accommodation. For the present, we acknowledge but set aside considerations of social perception to focus on the collective dynamics due to transmission across generations.

### The Computational Model

We develop an agent-based model of a population undergoing a vowel shift. The population consists of 10,000 agents. Each agent is initially assigned an age between 2 and 70 years. For simplicity, we assume that, as the simulation progresses and the agents' ages increase, agents past the age of 70 die and are replaced with an equal number of agents of age 2, thus keeping the population size constant over time.

New agents begin at age 2 because we abstract away the problem of initial vowel system acquisition, since we are primarily interested in vowel shifting. Vowel system acquisition by infants is non-trivial and has been studied through computational modeling by de Boer (2003). He showed that infants can acquire the vowel system of their parents through a combination of careful articulation by their parents (in child-directed speech), and compensatory expansion of articulations of reduced speech sounds (by the infants). We do not include these complexities in our model, assuming that some such mechanism is present to allow children to acquire the vowel system of their parents in the first two years of their life.

In our model, each new agent is assigned the vowel system of a randomly chosen parent agent. The parent agent is chosen from the subset of the population aged between 21 and 31 years. Adaptation of the vowel system happens between the ages of 2 and 16 years (inclusive). Once an agent reaches 17 years of age, its vowel system becomes fixed.

One simulated year consists of 5,000,000 interactions, where each idealized interaction consists of a *speaker* communicating the position of one of its vowels to a *hearer*. The speaker is chosen from the population a year older than the hearer population. The hearer updates the position of its own corresponding vowel in response, following a sys-

tem of internal constraints explained below. Since agent ages vary from 2 through 70, but learning only happens between the ages of 2 through 16, we have approximately  $15/69 * 10000 \approx 2174$  learning agents on average in the population. Since the vowel system consists of 6 vowels, this results in  $5000000 / (2174 * 6) \approx 383$  updates per vowel per agent per year, which is close to one update per vowel per agent per day.

Updates happen through *accommodation*, which simply means that the hearer tries to move the position of its vowel closer to the perceived position of the speaker's vowel.

The key assumption of our model is that accommodation incorporates representational momentum (Kelly and Freyd, 1987; Freyd et al., 1990). The principle of representational momentum is well-studied in psychology, and states that humans have a forward memory asymmetry for pitch (in the auditory case; it has also been demonstrated for other sensory modalities). When subjects are presented with a sequence of tones of rising pitch, they later recall the pitch of the final tone to be higher than it was. Freyd and her colleagues have also shown that the distance between the remembered final pitch and the actual final pitch is proportional to the implied velocity. Therefore they explain this phenomenon in terms of a momentum effect.

The notion of momentum is also common in artificial neural network learning, where it is used to improve convergence time and reduce oscillations in the weights (e.g., see Haykin, 1998, p. 170).

In the model we include a momentum term in the vowel update as,

$$v_i^{t+1} = v_i^t + (1 - \alpha)\eta(v'_i - v_i^t) + \alpha(v_i^t - v_i^{t-1}), \quad (2)$$

where  $\alpha$  is the momentum factor,  $\eta$  is the learning rate,  $v'_i$  is the target position for vowel  $v_i$ , and  $t$  is the time step.

After the new position of the vowel has been calculated in this way, two constraints are applied to decide its final position in time step  $t + 1$ .

- *Differentiability constraint*: If the speaker's vowel position is too close to the position of an alternate vowel in the hearer's vowel system, the hearer will not accommodate.
- *Margin of security constraint*: If an update brings a vowel too close to another vowel, both vowels get pushed apart.

If two vowels get too close to each other, there is a chance they will *merge*. The differentiability constraint acts to prevent mergers between vowels in the hearer's vowel space, and the margin of security constraint acts to repair the system if two vowels get too close. In reality, mergers do occur, and the relation between mergers and shifts is poorly understood. If mergers occur, language users can rely on other cues such as vowel duration or conversational context to disambiguate meaning. We don't model these aspects here.

The implementation of the differentiability constraint is straightforward. If  $v_1$  is the vowel being communicated, the hearer compares the perceived position of the speaker's  $v_1$  with the positions of the hearer's vowels  $v_2$  through  $v_6$  (i.e., all vowels other than  $v_1$ ). If the distance to any of these vowels is too small, the hearer does not update its  $v_1$  position, i.e., we just reset  $v_1^{t+1}$  to be the same as  $v_1^t$ .

The implementation of the margin of security constraint is a little more complicated because we have to decide a direction for each vowel to move when they push each other. For this we rely on a set of principles distilled by Labov.

### Labov's principles

Based on a survey of chain shifts in English and other languages, Labov (1994) proposes that there are three universal principles that constrain the possible movements of vowels.

- Principle I: In chain shifts, tense nuclei ("long" vowels, e.g., BOOT, BEET) rise along a peripheral track (regions close the  $f1/f2$  axes).
- Principle II: In chain shifts, lax nuclei ("short" vowels, e.g., BIT, BET) fall along a nonperipheral track.
- Principle III': In chain shifts, tense vowels move to the front along peripheral paths, and lax vowels move to the back along nonperipheral paths. (Note that this is numbered III' because he revised his earlier principle III).

He has applied these principles to the vowel movements in the NCVS as follows. Principle I applies to the fronting and raising of BAT. Principle II applies to the lowering of BIT (and to BET to some extent). Principle III' applies to the fronting of BOT and the backing of BET and BUT.

In our implementation of the margin of security constraint, therefore, if there is a collision (an update brings two vowels too close to each other), they move in the directions suggested by Labov's principles in an attempt to repair the vowel system. Note that the margin of security may not get re-established in a single update. There is no attempt to re-check after the update if the margin of security constraint is still being violated. This check effectively happens the next time the hearer agent again attempts to update the same vowel and notices the constraint violation again. The distance the vowel moves is determined by a push rate,  $\gamma$ , simply as

$$v_i^{t+1'} = v_i^{t+1} + \gamma u_i, \quad (3)$$

where  $u_i$  is a unit vector in the direction suggested by Labov's principles for vowel  $v_i$ .

### Experiments

As mentioned earlier, the population consists of 10,000 agents. We initialize 40% of the agents in a shifted state, i.e., with raised BAT, fronted BOT, and lowered BOUGHT. The rest of the population has the default initial state with low BAT, back BOT and mid-back BOUGHT. The positions

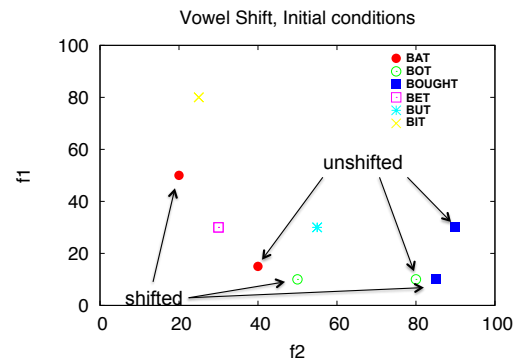


Figure 3: Initial conditions. 40% of the population has the shifted versions of the vowels BAT, BOT, and BOUGHT. The rest have the unshifted versions. The entire population has the same vowel positions for BET, BUT, and BIT.

of the other vowels are the same for all the agents. Both sets of vowels are shown in figure 3.

Parameter settings are as follows: learning rate,  $\eta = 0.0001$ , momentum,  $\alpha = 0.2$ , and push rate,  $\gamma = 0.00004$ . Vowel production is allowed to be noisy, by adding a random variable sampled from a circular Gaussian with zero mean and standard deviation 10.

Noise alone, even if momentum is zero, can cause a partial shift. This is an interesting finding in itself. So for comparison we also ran an experiment where all the settings are exactly the same, except that the momentum factor,  $\alpha$ , is set to 0.

We run the simulation for 160 simulated years, and then extract the average vowel positions for age-groups that are 20 years apart.

The results of the experiments, and a statistical comparison of the cases with and without momentum are presented in the next section.

### Results and Comparison

The results of the two experiments are plotted in figures 4a and 4b. The figures show the mean position of each vowel for each age groups. The age groups are chosen to be 20 years apart each, giving us eight groups from 160 simulated years. Group 1 corresponds to the subset of the population that was 20 years old at the beginning of the simulation, thus their vowel systems are in the initial condition and don't change. The next group corresponds to the subset that was 20 years old in year 20, followed by the group that was 20 years old in year 40, and so on. In each case, they are sampled once the agents are past the point of adapting their vowel systems.

From figure 4, it is immediately obvious that we see much more shifting in the case with momentum. A clearer comparison of the magnitude of shifting is seen in figure 5,



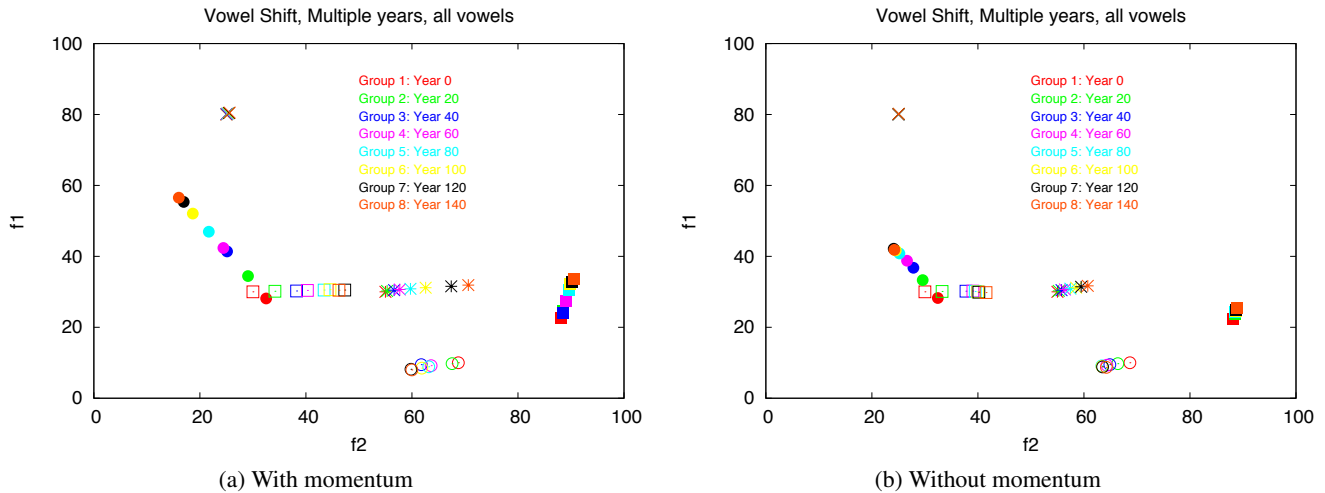


Figure 4: Vowel shifts generated by the model with and without representational momentum. For each age group, we plot the mean position (over that age group) of each vowel. The point shape denotes the vowel and the point color denotes the age group. We see that shifts are much more distinct when the model includes momentum.

which also shows a comparison with the data in figure 2. In each case, the magnitude of the shift is greater with momentum than without. Note that since BOT is fronting, i.e., moving in the direction of decreasing  $f_2$ , the plot in figure 5b shows that the shift with momentum is greater for BOT because the curve for BOT with momentum is below that without momentum.

The comparison with data is done by linearly transforming the empirical data so that the position for BOT from the speaker group born during 1890-1910 lies on top of the position for BOT from group 4 of the simulation with momentum. All the other data points are then transformed using the same mapping. Even this naïve mapping shows a good fit with the simulation results.

BAT and BOUGHT show close overlap between the empirical data and the simulation results with momentum. BUT shows a close match in the slope, i.e. the magnitude of the shift from one generation to the next, though the position suggests that we have chosen the initial position to be too fronted.

BOT shows a weaker match between data and simulation results, but as we will see below, in the case of that vowel, most of the shifts are not statistically significant. BET shows a steeper change in the data than in either simulation, though it is closer to the simulation with momentum.

The largest mismatch between the data and the simulation results for each vowel is for the earliest group, where the sample size is the smallest, consisting of only 3 speakers.

We do not show error bars in the plots to avoid clutter. However, we present a detailed statistical analysis below.

In order to examine change over time, we do a MANOVA to examine the statistical relationship between age group and

vowel position. The partial  $\eta^2$  values are shown in table 1. This is a measure of the effect size. It tells how much of the variation in the shift of a vowel is attributable to the age group. We see that the values for the experiment with momentum are higher than or close to the values for the experiment without momentum, and are closer to the data.

Table 1: A comparison of the partial  $\eta^2$  with and without momentum, and from the data. We see that the variation attributable to the group is in general larger for the experiment with momentum, and closer to the data.

Vowel	Partial $\eta^2$		
	w. m.	w.o. m.	data
BAT f1	0.276	0.19	0.34
BOT f2	0.045	0.036	0.226
BOUGHT f1	0.207	0.022	0.209
BET f2	0.195	0.207	0.558
BUT f2	0.503	0.559	0.346

As a rule of thumb, values of partial  $\eta^2$  of 0.1 or below are considered small, values around 0.2 are considered medium, and anything over 0.3 is considered a fairly big effect. The largest effects are seen for BUT, and for BAT with momentum. BOUGHT with momentum shows a significantly larger effect than without momentum.

We also do post-hoc tests between the age groups for each vowel, with and without momentum, and from the data. We don't have the space here to present the entire set of p-values, but the results can be summarized as follows.

- For the movement of BAT with momentum, all the pairwise tests are significant at the  $p < 0.05$  level except for groups 3 and 4, and groups 6, 7, and 8. Without momen-

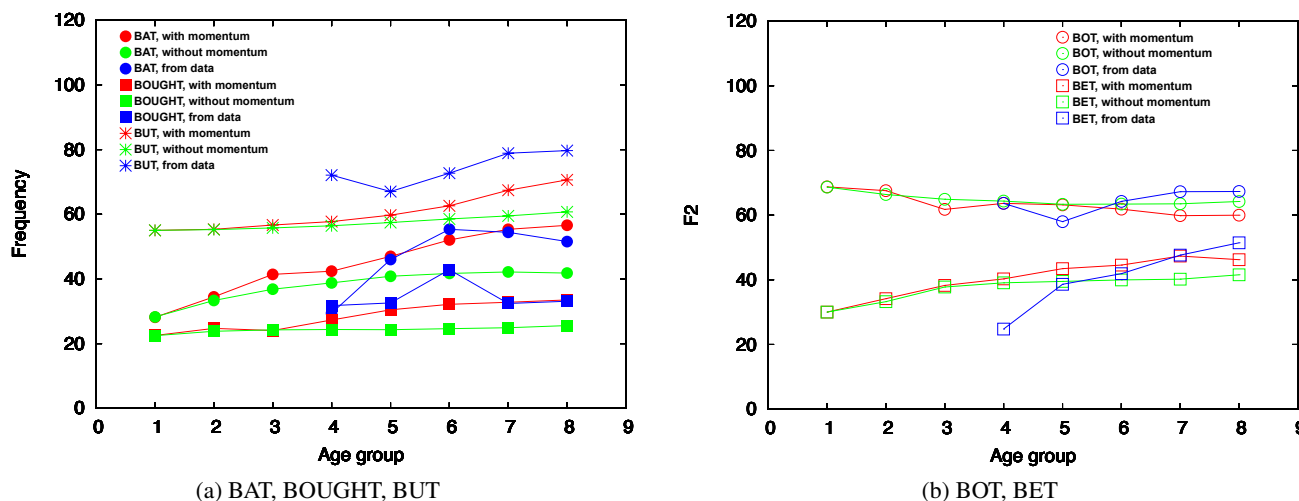


Figure 5: A comparison of the shift magnitude with and without momentum, and with the transformed empirical data. Only movement along the significant axis for each of the five vowels of interest is shown. The axes of interest are  $f_1$  for BAT and BOUGHT, and  $f_2$  for BOT, BET, and BUT. The comparison for BET is plotted along with BAT and BOUGHT to avoid overlap.

tum, the shifts are not significant from group 5 onwards. In the data, the shifts of BAT are only significant between the first group (1890-1910) and the rest.

- For the movement of BOT with momentum, the shifts are not significant from group 3 onwards, while without momentum, they are not significant from group 2 onwards. In the data, the shifts of BOT are not significant.
- For the movement of BOUGHT with momentum, the shifts are not significant early and late, i.e., between groups 1, 2, and 3, and from group 5 onwards. Most of the significant shifting happens from groups 3 to 4 to 5, though 5 vs. 8 is also significant ( $p < 0.005$ ). Without momentum, none of the movements are significant at  $p < 0.05$ , except 1 vs. 8. In the data, the shifts of BOUGHT are not significant.
- For the movement of BET with momentum, the shifts are not significant from group 5 onwards, while they lose significance from group 3 onwards without momentum. In the data, the shifts are significant when comparing groups 1 and 2 with the rest, but not amongst the rest.
- For the movement of BUT with and without momentum, the shifts remain significant throughout except for groups 7 vs. 8 in the case with momentum. In the data, the shift between group 2 and 5 is significant, but the others aren't.

## Discussion

To fully comprehend the results, we have to consider both the magnitude and the significance of the shifts. Taken together, several interesting inferences can be drawn from the results.

First, overall, shifts with momentum are larger, and continue for longer, than shifts without momentum. Shifts with momentum naturally come to an end on a time-scale of about 100-140 years, as the overshoot due to momentum dies out when movements get smaller, which matches well with observations. The magnitudes of the shifts in the data match quite well with the simulation with momentum.

More specifically, we see that BAT and BOT start moving more or less simultaneously, but BOT completes its movement early, while the movement of BAT slows down between groups 3 and 4 but gets a boost due to the push interaction with BET. This suggests a possible resolution to the ordering controversy between BAT and BOT mentioned earlier. McCarthy (2009) has suggested, based on her data and contrary to the chronology suggested by Labov, that the ordering is not simply BAT fronting and raising followed by BOT fronting, but rather that the movements might be interleaved, with BAT fronting followed by BOT fronting followed by BAT raising. However, it was not known what might cause such interleaving.

Our simulation is supportive of McCarthy's chronology where we see BAT raising and fronting until it comes near BET, followed by BOT fronting, followed by the second stage of BAT raising and fronting due to the push interaction with BET. This suggests the push interaction as a possible cause for the interleaving.

The movement of BOUGHT with momentum is actually upward, contrary to Labov's suggestion, but matching McCarthy's data (fig. 2) where the position of BOUGHT is close to the initial position or even higher. Without momentum BOUGHT ends up in a position in-between the initial shifted and unshifted positions, which is unrealistic.

BET and BUT show strongly significant movement, both with and without momentum, but as can be seen from both fig. 4 and fig. 5, the magnitude of the shift is much greater with momentum. Without momentum, the groups bunch up very quickly, resulting in only a partial shift.

### Conclusion

We have presented an agent-based model of vowel shifts based on empirical principles that have been derived in the sociolinguistic community. We have shown that an additional ingredient, viz. representational momentum, is required to explain several aspects of the NCVS in Chicago.

The main missing piece is an account of the movement of BIT, which shows robust movement in Chicago but relatively little in our model. The data in fig. 2 suggests a possible interaction between BAT and BIT, in that BAT raising may push BIT backward in accordance with Principle III'. Alternatively, BIT's movement could be due to its involvement in a parallel shift, in which front lax vowels BET and BIT move backward as a class. We have not implemented this principle, although it may be worth considering.

More generally, we need a deeper account of how vowels move when they come too close to each other. Labov's principles are essentially *ad hoc* rules derived from observation. A cognitive and psychoacoustic perspective might provide an energy-function based approach to vowel spacing and interaction. Another approach might be based on neural coding, similar to Joanisse and Seidenberg (1997).

Another direction in which this work can be extended is to incorporate social and economic demographic-based variation, which might help explain the regional variations in the NCVS and in other shifts.

In conclusion, we believe that computational simulation has much to offer the study of vowel shifts and other large-scale dynamical phenomena in language because through simulation we can shed light on precisely those questions for which the available data are sparse and hard to gather.

### Acknowledgements

S. S. was supported in part by DTRA R&D Grant HDTRA1-09-1-0017, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF NetSE Grant CNS-1011769, NSF HSD Grant SES-0729441 and NSF PetaApps Grant OCI-0904844. C. M. was supported in part by the GMU College of Humanities and Social Sciences.

### References

- Babel, M. (2009). *Phonetic and Social Selectivity in Phonetic Accommodation*. PhD thesis, Dept. of Linguistics, University of California, Berkeley.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59:1–26.
- de Boer, B. (2001). *The Origins of Vowel Systems*. Oxford University Press.
- de Boer, B. (2003). Conditions for stable vowel systems in a population. In *ECAL03*, pages 415–424.
- Dras, M. and Harrison, K. D. (2003). Emergent behavior in phonological pattern change. In Standish, Abbass, and Bedau, editors, *Proceedings of the 8th International Conference on Artificial Life*, pages 390–393. MIT Press.
- Ettlinger, M. (2007). An exemplar-based model of chain shifts. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 685–688, Saarbrücken, Germany.
- Freyd, J. J., Kelly, M. H., and DeKay, M. L. (1990). Representational momentum in memory for pitch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6):1107–1117.
- Giles, H. and Coupland, N. (1991). *Language: Contexts and Consequences*. Open University Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2):251–279.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation (2<sup>nd</sup> ed.)*. Prentice Hall.
- Herndobler, R. (1977). *White Working-Class Speech: The East Side of Chicago*. PhD thesis, University of Chicago.
- Joanisse, M. F. and Seidenberg, M. S. (1997). [i e a u] and sometimes [o]: Perceptual and computational constraints on vowel inventories. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*.
- Kelly, M. H. and Freyd, J. J. (1987). Explorations of representational momentum. *Cognitive Psychology*, 19:369–401.
- Labov, W. (1994). *Principles of Linguistic Change Vol. 1: Internal Factors*. Blackwell, Cambridge, MA.
- Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Mouton de Gruyter, Berlin.
- Lakkaraju, K., Swarup, S., and Gasser, L. (2012). Consensus under constraints: Modeling the Great English Vowel Shift. In *The International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP)*, College Park, MD.
- Martinet, A. (1955). *Economie des changements phonétiques*. Francke, Bern, Switzerland.
- McCarthy, C. (2009). The Northern Cities Shift in real time: Evidence from Chicago. *Penn Working Papers in Linguistics*, 15(2):article 12.
- McCarthy, C. (2011). The Northern Cities Shift in Chicago. *Journal of English Linguistics*, 39(2):166–187.
- Stanford, J. N. and Kenny, L. A. (2012). An agent-based computational interpretation of the transmission and diffusion of vowel chain shifts across large communities. *Unpublished manuscript*.
- Thomas, E. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. Duke University Press, Durham, NC.
- Thomas, E. R. and Kendall, T. (2007). NORM: The vowel normalization and plotting suite. Online Resource: <http://ncslaap.lib.ncsu.edu/tools/norm/>.
- Trudgill, P. (1986). *Dialects in Contact*. Basil Blackwell, Oxford.