

The *Gettysburg Corpus*: Testing the Proposition that All Tense /æ/s Are Created Equal

Abstract

Corpus studies of regional variation using raw data from the internet focus predominantly on lexical variables in written language. However, online repositories such as YouTube offer the possibility of investigating regional patterns using phonological variables, as well. This paper demonstrates the viability of constructing a naturalistic speech corpus for sociophonetic research by analyzing hundreds of recitations of Abraham Lincoln's "Gettysburg Address." We first replicate a known result of phonetic research, namely that English vowels are longer in duration before voiced obstruents than before voiceless ones. We then compare /æ/-tensing in recitations from the Inland North and New York City dialect regions. Results indicate that there are significant regional differences in the formant trajectory of the vowel, even in identical phonetic environments (e.g., before nasal codas). This calls into question the uniformity of "/æ/-tensing" as a cross-dialectal phenomenon in American English. We contend that the analysis of spoken data from online social media can and should supplement traditional methods in social dialectology to generate new hypotheses about socially conditioned variation.

Keywords

short-*a* systems, Inland North, New York City, corpus sociolinguistics, YouTube

1. Introduction

The recent turn to social media sites as sources of raw data for sociolinguistic research has provided new opportunities for the analysis of regional variation in American English. In studies of synchronic variation, Twitter corpora have been used to replicate some of the core findings of traditional dialectological research, including the locations of major dialect boundaries (Huang et al. 2016). In studies of diachronic variation, geotagged tweets have also been used to examine the geographic diffusion of linguistic change, demonstrating that innovative lexical items spread most rapidly between communities with similar racial demographics (Eisenstein et al. 2014). While such large-scale corpus studies have proven effective in research on regional variation in *written* English, comparable work on variation in *spoken* English (e.g., from YouTube videos) has lagged behind. The focus on written texts in social media studies has understandably resulted in a bias towards lexical variables. While this allows for direct comparison to a number of early dialectological studies that were also largely concerned with vocabulary items (e.g., Kurath 1949; Cassidy 1985), it deemphasizes the phonological variables that were the focus of the *Atlas of North American English* (Labov, Ash, and Boberg 2006) and other studies of regionally differentiated sound systems using spoken corpora designed specifically for research in linguistics, such as the Philadelphia Neighborhood Corpus (see Labov, Rosenfelder, and Fruehwald 2013).

The relatively slow adoption of “found data” from the internet in sociophonetics can be attributed to the number of serious methodological and theoretical challenges they pose. Unlike text, which can be mined and analyzed using existing software packages and toolkits, spoken data is often locked up in video files that cannot be analyzed until the recordings have been

transcribed. Unfortunately, fully transcribed video files comprise only a small percentage of publicly available recordings (e.g., TED talks), and the output of automatic speech transcription (e.g., YouTube's automatic captioning) generally requires retranscription by hand before it can be used effectively in research. Left unchecked, such errors in transcription will result in increased unreliability further downstream, when those texts become the basis for phonetic transcription, forced alignment, and ultimately the extraction of phonetic measurements.

The lack of transcriptions is not the only reason for the slow adoption of social media data in sociophonetic research. Unlike spoken corpora compiled by linguists for their own projects, those built from online platforms like YouTube must grapple with inherently imbalanced user bases. Whereas sociolinguistic fieldwork can yield a demographically balanced sample of speakers, it is considerably more difficult to obtain demographic information from YouTube users; for example, information on a speaker's gender, age, race, geographic location, and language background can often be inferred only crudely and impressionistically, if at all. Depending on the social variables that are of interest to the researcher, publicly available videos on the internet may ultimately be inadequate as a source of speech data. Where demographic information is explicitly available, it must be taken at face value for the purposes of sampling (at least initially; see our discussion of post-hoc speaker selection in section 3.2).

Another point of contrast between corpora produced for traditional sociolinguistic research and those extracted from social media relates to recording-level environmental factors. Sociolinguistic interviews are typically conducted in a quiet room and digitally recorded, without file compression, using a uniform set of recording equipment. By contrast, videos from large online repositories vary considerably in environmental factors like background noise and camera

quality; a corpus of recordings built from YouTube, for example, may include videos recorded with laptop webcams or mobile phones in noisy coffee shops as well as using professional cameras and microphones in quiet studios. Because social factors like age and socioeconomic class also correlate with access to technology, this variation in recording equipment may confound genuine patterns of sociolinguistic variation. To make matters more complicated, YouTube's proprietary compression algorithms have raised questions about the reliability of its videos for acoustic analysis (De Decker and Nycz 2011; see also Bulgin, De Decker, and Nycz 2010). To some extent, one can mitigate the lack of uniformity in recording equipment by choosing a large and diverse set of recordings. One can also mitigate the effects of differing degrees of file compression by including by-recording or by-uploader random effects in statistical models. Nevertheless, these factors introduce a level of noise, in both the auditory and statistical senses of the word, that one should weigh carefully when assessing the validity and reliability of any acoustic analysis.

The goal of this study is to demonstrate the potential of large speech corpora compiled from online social media for automated sociophonetic analysis, and for dialectological research in particular. In order to resolve some of the challenges outlined above, we have opted to work backwards, first choosing a data source and then formulating a research question that the data are particularly well suited to address. Our data come from hundreds of recitations of Abraham Lincoln's "Gettysburg Address" (1863), which were solicited to celebrate the release of Ken Burns's 2014 PBS documentary *The Address*. The recitations were video-recorded by individuals from all fifty states, representing a variety of age groups and ethnic backgrounds; they are catalogued by state at LearnTheAddress.org, and the video files themselves are hosted on

YouTube. These recordings are especially appropriate for phonetic research because they all share the same transcript—the text of Lincoln’s speech—which greatly facilitates the automatic segmentation of sound files, even if it admittedly limits us to the exact number and variety of words included in the original speech. Because acoustic measurements can be obtained from the same lexical items across recordings, the researcher can also inherently control for factors such as phonetic environment, token frequency, and (to some extent) prosodic context, in much the same way as during a laboratory reading task. This creates a viable testing ground for hypotheses about the social factors that condition patterns of linguistic variation—in this case, the role of state residency.

In order to demonstrate the suitability of the *Gettysburg Corpus* for acoustic analysis, we first replicate the well-known finding in the phonetic literature that vowels are longer in duration before voiced obstruents than before voiceless ones (e.g., Chen 1970; Hooper 1977). We then use the recitation data from the *Gettysburg Corpus* to carry out a novel comparative analysis of /æ/-tensing in two regional varieties of American English. Our study reveals that speakers from Michigan and New York State who exhibit features of the Inland North and New York City dialects, respectively, differ significantly from one another in the phonetic implementation of /æ/-tensing. The results call into question whether “/æ/-tensing” (or “/æh/”) should be treated as a unified cross-dialectal phenomenon, even in identical triggering environments such as pre-nasal contexts.

What follows is a case study in the analysis of publicly available recordings for dialectological research. Although our focus is primarily methodological, since the research question was informed by the selection of an online data source and not vice versa, the

organization of our paper follows that of a more traditional sociophonetic study. First we provide an overview of the phonetic variable under consideration (/æ/-tensing) and summarize previous research on the systematic cross-dialectal differences in its conditioning factors. We then describe the methods of automated data extraction from YouTube recordings, the criteria used to select dialect speakers from the larger population of Michiganders and New Yorkers, and the measurements that entered into the statistical analysis. We then summarize the core findings of our study, including the replication of the well-known effect of contextual conditions on vowel length, as well as new findings about the different trajectories of /æ/-tensing in these two dialects. Finally, we discuss the implications of using social media to address topics in sociophonetic variation, arguing that despite their limitations, spoken corpora not specifically designed with linguists in mind can nevertheless offer a rich source of data for sociophonetic and dialectological research.

2. /æ/ in American English

The acoustic analysis of the short-*a* vowel (/æ/) is an ideal testing ground for a new methodology because short-*a* has received extensive coverage in the dialectological and sociolinguistic literature (e.g., Trager 1930; Cohen 1970; Callary 1975; Labov 1989; Goodheart 2004; Becker and Wong 2010). It is well known that variation in the production of the vowel is subject to both phonetic and phonological conditioning. Studies utilizing a static measurement taken at a single point in time find two distinct variants: one in low front position and one that is comparatively raised and fronted (Labov, Ash, and Boberg 2006, 173). Which variant surfaces is phonologically conditioned; however, the phonological conditioning varies from region to

region. The best known of the attested patterns of short-*a* systems are the following (Labov, Ash, and Boberg 2006, 173–84):

1. the nasal system, in which vocalic nuclei are high and front before nasals, in both closed and open syllables (the most common pattern across the US);
2. the “short-*a* split” of New York City, which involves raising and fronting only before voiceless fricatives (*half*), voiced stops (*grab*), and the front nasals /n/ and /m/ (*ham*) in closed syllables;
3. the “short-*a* split” of Philadelphia, which involves raising and fronting only before front voiceless fricatives, a limited number of voiced stops (e.g., *mad*, *bad*, *glad*), and the front nasals in closed syllables; and
4. the Inland North dialect, spoken in cities such as Chicago, Detroit, and Rochester, which involves raising and fronting in all environments.

Note that the Inland North pattern, unlike other patterns, does not involve allophony. Rather, the raised and fronted variant always surfaces. This does not mean that production of /æ/ is invariant; rather, the following consonant conditions the degree of raising/fronting. Typically, the vowel is most raised and fronted in pre-nasal contexts and lowest and backest in pre-dorsal contexts. It is additionally raised/fronted when preceding voiced obstruents versus voiceless ones (Callary 1975; Goodheart 2004). The differences between this phonetic conditioning and the phonological conditioning described above are that the differences in production in the Inland North system are smaller and that there is still a considerable degree of overlap between environments.

The distinctions between these systems are often complicated when one examines the behavior of particular individuals and communities. For example, in recent work on the Philadelphia dialect, Labov et al. (2016) describe competition between the city's "traditional" (split) system and the nasal system, with the latter replacing the former among speakers engaged in the pursuit of higher education.

Although the majority of acoustic analyses involve single point measurement, it is well known that there is variation in the vowel dynamics, as well. Jacewicz and Fox (2013), for example, show that /æ/ varies with respect to vowel inherent spectral change (VISC) between the Midland, South, and Inland North. The *Atlas of North American English* (Labov, Ash, and Boberg 2006, 176–8) identifies one aspect of the vowel trajectory that appears to be geographically restricted to the Inland North. Here, a trajectory that they term "Northern breaking," in which the vowel is a diphthong ([ɛə, ɪə], etc.), can surface. The authors describe this as produced with two steady states. The first half of the vowel is raised and fronted. Because it begins in mid or high position, it then transitions to a second state by backing and lowering to reach a more central position in the second half. This is in contrast to a more common vowel trajectory found in cities such as New York and Philadelphia, in which a raised and fronted [æ] develops a slight peak in F2 (fronts slightly) and resolves in an offglide by backing to a more central position. Labov, Ash, and Boberg (2006) term this offglide an "inglide." The vowels that display Northern breaking tend to be considerably longer in duration than both the inglided, raised and fronted variant and the traditional low front variant (Labov, Ash, and Boberg 2006, 177).

While these observations are certainly suggestive of geographic differences in the

realization of the raised and fronted variant, it is unknown how robust these patterns really are. It is not known, for example, whether Northern breaking is found in all environments that trigger raising and fronting, or only in some. Fox and Jacewicz (2009) suggest that there may be variation based on context, as they find that for a set of speakers from Wisconsin, formant movement is greater when preceding voiced stops than when preceding voiceless stops. However, studies have not explored a wider range of phonological contexts. Similarly, it is also not known how common the Northern breaking pattern is in the North, as a proportion of all raised and fronted tokens produced by Northerners. Labov, Ash, and Boberg (2006) indicate that the pattern is rarely found in dialects beyond the North (hence its name), but also that individuals in the North themselves vary in their production of raised and fronted tokens. In fact, their example of Northern breaking (p. 177, figure 12) and their example of ingliding (p. 177, figure 11) were both produced by the very same speaker, Sharon K. of Rochester, New York. While the authors present some data indicating that speakers from the Inland North produce more tokens with Northern breaking than do speakers from other regions (Labov, Ash, and Boberg 2006, 178), they do not state explicitly how these findings were obtained. However, their discussion suggests that auditory coding and reading of spectrograms were used rather than a multi-point measurement. This question of differences in formant movement across dialects does not appear to have been explored elsewhere in the literature. While Fox and Jacewicz (2009) show a difference in formant movement between speakers in Wisconsin, Ohio, and North Carolina, these differences derive from a comparison of a set of speakers who tense in the environments tested and two sets who do not.

Sociolinguists and phonologists seem to suggest that the raised and fronted /æ/ variant is

realized uniformly across the phonetic environments in which it is found, for example, before nasals vs. before voiceless fricatives (e.g., Labov, Ash, and Boberg's [2006] use of "/aeh/" for all tense environments in New York City; Duncan's [2016] use of /εə/ to approximate a diphthong for all tense environments in the Northern Cities). While the raised and fronted variant may seem uniform when /æ/ is measured at a single point in time (such as the F2 peak), there may still be contextual differences in the formant trajectory over the duration of the vowel, beyond those that can be attributed to transitions from/to the preceding/following segments. To our knowledge, no dialectological studies have pursued this question with a sample that compares tensing dialects.

The raised and fronted variant is conventionally described as tense, in opposition to the historically lax vowel. There are differing reasons for the claim depending on the pattern in question. For the short-*a* split of New York and Philadelphia, one claim is that there is a phonemic rather than allophonic split (Labov, Ash, and Boberg 2006, 175). Because the raised and fronted variant in these regions has an inglide, the argument is that the phoneme is thus like other tense vowels that exhibit glides. With respect to the Inland North pattern, the claim is twofold. The first claim is that tense vowels rotate clockwise around the vowel space in chain shifts (Labov 1994, 213). Because the Inland North vowel participates in the Northern Cities Shift, it therefore must be tense. The second claim is that the diphthongal Northern breaking variant is bimoraic (Labov, Ash, and Boberg 2006, 177). It is not clear how the authors define the presence of two moras, but at the very least the term seems to refer to the vowel's relatively longer duration: Labov, Ash, and Boberg suggest that the two states of the diphthong are each long enough to constitute short vowels if isolated. However, given claims in phonology that in English tense vowels are bimoraic and lax vowels are monomoraic (Green 2001), this suggestion

would require that Inland North /æ/ be tense.

The claim that the raised and fronted variant is tense therefore represents a claim with respect to the phonological and articulatory status of the vowel. Such a claim is more tenuous than the conventionalized label “tense” would suggest. For example, research in experimental phonology regarding the status of /æ/ among Inland North speakers is inconclusive. Duncan (2016) conducted a forced-choice nonce word task which tested whether participants accepted the vowel in environments that only lax vowels may appear in. The results suggest that the vowel remains part of the lax natural class. In contrast, Nesbitt (2018) conducted a syllabification task which tested whether participants treat syllables with /æ/ like syllables with tense vowels or like syllables with lax vowels. The results indicate that for older speakers, the vowel patterns with the tense natural class. It is unclear whether the differences in results are due to differences in the populations sampled (perhaps different groups behave differently) or to the task itself (the vowel is treated differently depending on environment).

Articulatory studies conducted under controlled laboratory conditions have also found significant differences within and across individual speakers. Articulatory tenseness suggests that speakers produce the vowel with an advanced tongue root. A distinct formulation of a claim to tenseness could also mean that the backing/lowering of the inglide involves tongue retraction over the course of production. Relying on acoustic measurements as well as an ultrasound analysis of the tongue in a controlled reading task, De Decker and Nycz (2012) have shown that the mapping of articulatory tenseness (advancement of the tongue root and raising of the tongue body) onto acoustic tenseness (higher F2 and lower F1 frequencies, respectively) can vary considerably from speaker to speaker, word to word, and even token to token. In their study, the

two speakers who have acoustically raised and fronted vowels only before nasals differ markedly from one another in their articulations: one speaker exhibits a raised and fronted lingual shape, while the other does not, and instead appears to rely primarily upon coarticulatory nasalization. The authors conclude that future work could shed light on whether (and how) these individual differences map onto broader regionally or socially conditioned community patterns.

It is ultimately unclear, then, whether the raised and fronted variant is indeed tense. While we will keep the classification of variants as tense/lax moving forward, this is done more as a convention than as a claim about phonology or articulation. The key point is that while the tense variant is often discussed as roughly the same across dialects, some evidence suggests that this may not be the case when considering the formant trajectory of the vowel. Furthermore, there is little or no work that examines whether the tense variant is the same across phonetic contexts. As such, our goals in applying our novel methodology for automated sociophonetic analysis are to ascertain whether the tense variant has the same formant trajectory in two dialects, and whether the formant trajectory within dialects differs across phonetic contexts.

3. Methods

3.1. The *Gettysburg Corpus*

The data analyzed in the current study come from video-recorded recitations of Abraham Lincoln's "Gettysburg Address," compiled at LearnTheAddress.org and organized according to the reciters' self-reported state affiliations. A script was used to scrape (i.e., automatically extract) the YouTube link to every video included on every page for both Michigan and New York, and a second script was used to download a .wav file containing the audio track of each

video file.¹ These .wav files were saved to a hard drive and organized into directories corresponding to state groups. In total, 402 audio recordings were downloaded: 230 from Michigan and 172 from New York.²

In order to generate a transcript file for each recording that would comply with the formatting specifications of the Forced Alignment and Vowel Extraction (FAVE) program suite (Rosenfelder et al. 2014), a Python script was authored to produce a text file containing each speaker's identification code (i.e., the unique ID of their YouTube video), the start and end time of their recording, and finally the text of the "Gettysburg Address," entered as a single breath group. Several versions of Lincoln's speech are publicly available; because LearnTheAddress.org distributes the so-called "Bliss copy," we use this version as the text in our transcript files. The paired .wav recordings and transcript files were then forced aligned in FAVE, which generated textgrids that can be analyzed acoustically. In keeping with our goal of testing whether "found data" such as this can be used for *automated* analysis, we did not systematically hand-check the vowel boundaries output by the forced aligner. Rather than extract vowel formants using FAVE, we wrote a script in Praat (Boersma and Weenink 2017) to take measurements of F1 and F2 frequencies (in Bark) for all stressed tokens of /æ/ at the maximum of F2. We also measured F1, F2, and F3 for these tokens and a handful of other stressed vowels (see below) at 10% intervals across the vowel duration, for a total of nine measurements per vowel. All measurements are based on the setting nFormants=5 in Praat.

The matter of which lexical items to include in the analysis is highly constrained. This is one cost of an opportunity sample, as we are limited by the transcript of the "Gettysburg Address." The speech is famously terse (269 words), and therefore there are relatively few total

tokens of *any* variable within it, let alone tokens of a variable in a specific environment. For the analysis of /æ/, we focus on four particular lexical items—*add*, *advanced*, *last*, and *task*—which represent phonetic environments that typically trigger tensing in both the Inland North and the New York City dialects. Whether *add* consistently triggers tensing in the New York City dialect is a matter of debate. According to Labov (2007), word-initial environments generally do not trigger tensing, regardless of the following environment. However, Coggshall (2017) shows that many speakers of the New York City dialect from Jersey City (New York’s “Sixth Borough”) either tense or variably tense in this environment. We follow Coggshall’s findings by treating *add* as representing a phonetic environment that triggers tensing. Our analysis will thus include four tokens of /æ/ per speaker, which occur in different environments. While we acknowledge that it would be ideal to have more tokens per environment and more tokens per speaker, this is unfortunately precluded by the data source.

We report F1/F2 values using Thomas and Kendall’s (2007) modification to the Bark Difference Metric (Syrdal and Gopal 1986). This method is particularly appropriate for our data because it is vowel-intrinsic, applicable without knowing speaker sex, and can be applied to a small number of tokens. Although it stretches female speakers’ vowel space in the F1 dimension (Clopper 2009), this is less problematic for our purposes because we are not considering high vowels in this study. It obtains a normalized F1 through subtracting F1 from F3, and a normalized F2 through subtracting F2 from F3. While we follow the method outlined in the online NORM suite, we self-implement this normalization in R (R Core Team 2017).

The limited number of tokens per speaker and per environment is in fact but one of the many ways in which our data source is admittedly noisy. To reiterate, while a goal of this study is

to investigate the formant trajectory of tense /æ/, our primary goal is to test whether it is possible to use spoken language data found online for automated sociophonetic analysis. There are several other sources of noise in the data, such as the fact that speakers recorded themselves both indoors and outdoors in spaces with varying acoustics; that speakers recorded themselves using a range of camera and microphone types, some of which may have made higher quality recordings than others; that the speakers are of an unknown variety of social backgrounds, including gender, age, etc.; that some speakers treated the speech event as a performance while others merely read the text aloud; and that it is unknown how exactly YouTube's proprietary compression algorithm works.³ These sources of noise would all be controlled for in a traditional study but cannot be controlled for in this one. In some ways, the central research question here is thus whether any patterned variation can be pulled out of this noise. In our favor is that the noise in the data is likely to be independent of speakers' state residency; whatever issues are present in the Michigan data will also be present in the New York data and therefore should not confound otherwise systematic state-level phonetic differences. Moreover, while it is not possible to control for the variable compression of video files as users uploaded them to YouTube, the files from both states were all downloaded at the same time and at the maximum allowable quality, using a desktop computer with a wired internet connection. Nevertheless, these factors mean that specific formant values should not be taken at face value, and indeed we emphasize relative differences in our analysis. Our use of the *Gettysburg Corpus* is intended as a methodological proof of concept in which we examine a small-scale corpus with an eye toward larger studies. Hence, we trade the quantity of tokens for ease of automation. Using another corpus of found data in which speakers do not follow the same transcript may yield more tokens of a variable, but would also

require additional time and labor to process.

3.2. Selection of Speakers

In order to test whether speakers of the Inland North dialect and speakers of the New York City dialect have different phonetic implementations of /æ/-tensing, it is necessary to remove speakers who do not exhibit features of the target dialects. In other words, given the messy nature of our data set, it is to be expected that some subset of the individuals who submitted their readings of the “Gettysburg Address” and indicated that they live in Michigan or New York State are actually not speakers of the desired varieties. One reason for this is that a number of the recordings in both states were produced by local colleges or television news channels, and may have included individuals visiting from other parts of the country or from abroad. (At the same time, these more professional productions also allowed individuals who might not have YouTube accounts of their own, including older Americans, to participate in the “Learn the Address” project.) Given that recitations of the “Gettysburg Address” are a relatively formal or “careful” speech event, much like other reading tasks, it is also inevitable that some number of speakers will adjust their pronunciations towards a perceived standard befitting this kind of performance. To the extent possible, our goal was to include only those speakers whose recitations exhibit the tensing patterns that have been described for the two dialects, i.e., the general /æ/-tensing pattern for the Inland North dialect and the short-*a* split for New York City.

This goal is also complicated by the fact that dialect boundaries do not necessarily match state boundaries. This is true for Michigan; not all speakers from the state speak the Inland North dialect, and parts of the state like the Upper Peninsula have been described as having dialect features quite different from those of the Inland North (e.g., Rankinen 2018). It is especially true

for New York State, which includes speakers of several dialects: New York City English in New York City and parts of Long Island (Labov 1966; Becker 2010; Newman 2014; Shapp 2019), the Inland North in cities like Syracuse, Rochester, and Buffalo (Labov, Ash, and Boberg 2006; Driscoll and Lape 2015; Milholland 2018), Western New England in other parts of the state, and transitional areas among the dialects (see Dinkin 2013 for discussion). This means that identifying speakers with the general /æ/-tensing pattern for the Inland North dialect and the short-*a* split for New York City requires targeting a subset of speakers from these states.

This is further complicated by the well-described retreat from the Northern Cities Shift and traditional New York City dialect features. In the Inland North, the NCS has been found to be retreating in apparent time in many cities (Syracuse: Driscoll and Lape 2015; Buffalo: Milholland 2018; Chicago: D’Onofrio and Benheim 2019). This retreat has been especially well-documented in Michigan. In Lansing, for example, the apparent-time retreat from the NCS includes a reorganization of the /æ/ system away from the general tensing of the Inland North to the nasal system (Wagner et al. 2016); similar findings have been reported for Kent County in western Lower Michigan (Rankinen, Albin, and Neuhaus 2019). A similar reorganization is occurring in New York City, where the short-*a* split configuration is giving way to the nasal system in apparent time (Becker 2010; Haddican et al. 2018). These retreats mean that even if the YouTube videos that we collected were geotagged with a specific location (which they are not), we would still not be able to use such information to accurately obtain our subsets of Inland North speakers with general /æ/-tensing and New York City speakers with the short-*a* split.

Because one goal of this study is to determine whether a fully automated acoustic analysis of found data is feasible, we needed to implement a post-hoc approach to obtain the

desired subsets of speakers from the two states. That is, we define speakers' membership in a subset of "dialect speakers" based on their exhibiting acoustic measures fitting that subset, rather than any social characteristics beyond their listed state residency.

In our approach, we defined computer-implementable linguistic criteria for determining whether a speaker would be expected to exhibit the patterns of /æ/-tensing associated with the Inland North and New York City dialects. These criteria were designed to be relatively conservative; we aimed to err on the side of excluding /æ/-tensers rather than including non-/æ/-tensers. For Michiganders, we were able to define this solely by reference to the production of /æ/. Because we take Inland North /æ/ to be the generally raised and fronted vowel found in the NCS, we can define an Inland North speaker as one who has a small difference in production between pre-nasal and pre-oral contexts. If the difference is large, the speaker has a nasal system, and therefore by definition does not have NCS /æ/ and is not an Inland North speaker. Given this, we calculated the difference in F1 (on the Bark scale) measured at the maximum F2 of the words *add* and *advanced*, both of which are produced in the third paragraph of the "Gettysburg Address."⁴ If F1 of *advanced* was greater than F1 of *add* by at least 0.75 Bark units (about 100 Hz)—i.e., if the two vowels were of sufficiently different height—it could be inferred that such a speaker has a nasal system, such that /æ/ is tense before /n/ but not before /d/, rather than the Inland North pattern of tensing in both contexts. These speakers were removed from the data set. Furthermore, if F1 (Bark) measured at the maximum F2 of *add* was more than 6.5—more or less coinciding with the *Atlas*'s normalized use of 700 Hz to distinguish tense/lax tokens—the speaker was removed. Only those speakers who met both criteria, and therefore appear to engage in the general /æ/-tensing of the Inland North, were included in our analysis (n

= 112 / 230, or 48.7%). That so few speakers are included is likely a testament to the conservative nature of our criteria and to the fact that many Michiganders do not have general /æ/-tensing whether for reasons of regional variation or age-stratified retreat from the feature.

The selection of speakers from New York posed a much greater challenge for reasons of regional variation. Applying the same criteria used for Michigan would successfully include speakers who exhibit the split-*a* system of New York City. However, the criteria would also include speakers from cities located in the Inland North dialect region like Syracuse, Rochester, and Buffalo. The obvious solution would be to add a comparison to a lexical item that contained /æ/ in an environment in which the lax variant is produced by speakers with the short-*a* split. Because the NCS system would have the tense variant in such an environment, we would be able to distinguish between Inland North speakers and New York City dialect speakers. Unfortunately, this is a situation in which we encounter the limitations of the “Gettysburg Address”: no lexical items in the text cleanly provide this context in a stressed environment. Available lexical items include function words (several instances of *that*), those with vowels following a liquid which therefore risks coarticulation effects (*rather*, *detract*), or words that are archaic and not necessarily part of every speaker’s active vocabulary (*hallow*). The best candidate is *battlefield*, which should be lax by virtue of /æ/ preceding a (phonemic) voiceless stop and being in an open syllable. Both of these are contexts that yield the lax variant in theory (and the lexical item likely does contain a lax /æ/ in New York City English; Allison Shapp, p.c.). However, that it precedes a flap makes this environment less than ideal. In the interest of a maximally cautious approach we thus appear to require an /æ/-extrinsic method of inferring New York City dialect speakers.

Our solution is to adopt a more holistic approach that assumes sociolectal coherence

(Guy 2013): that the prevalence of linguistic features is correlated within a dialect, meaning that individual speakers would produce variants of multiple dialect features at similar rates. In his initial exploration of coherence in Brazilian Portuguese, Guy finds limited evidence of coherence based on social factors and calls for further research into whether it consistently occurs.

Fortunately for us, Becker (2016) explores whether the New York City dialect displays sociolectal coherence. Within her sample, she finds that three features—non-rhoticity, the short-*a* split, and raised /ɔ/—are highly correlated in use across the community as a whole. She finds less evidence of coherence on the part of individual speakers, although older white speakers do tend to show coherence. Each of the three features that Becker examined are well-known features of the New York City dialect (Labov, Ash, and Boberg 2006; Becker 2010; Newman 2014), and each is in retreat (Becker 2010; see also Wong 2012). This means that the speakers who do display coherence are likely to be the speakers that we are looking for, because the speakers displaying coherence are the ones with traditional New York City dialect features. This alone would not be rigorous enough for our purposes, however. Crucially, Becker (2016) finds an implicational hierarchy with respect to sociolectal coherence: while a speaker with a raised /ɔ/ can have either the short-*a* split or the nasal /æ/ system, the inverse is not true. Becker finds that no speaker in her sample with the traditional short-*a* split of the New York City dialect has a lowered /ɔ/; all have the raised variant.

This finding allows us to operationalize sociolectal coherence to obtain a subset of New York City dialect speakers from New York State. We maintain the criteria used to obtain the subset of Inland North speakers from Michigan. To this we add an additional criterion: that speakers have a raised /ɔ/. Because raised /ɔ/ is not found in the Inland North dialect, if speakers

who display /æ/-tensing also pass this criterion, Becker's (2016) results suggest that we may be confident that the speakers are New York City dialect speakers with the short-*a* split. Our use of this criterion should not be taken as a claim that there is a *structural* link between raised /ɔ/ and the short-*a* split; that speakers with a nasal system can nevertheless have a raised /ɔ/ in fact appears to be evidence to the contrary. We do believe, however, that there is sufficient evidence of a *social* link between the variables for the criterion to prove useful. We calculated the difference in F1 (Bark) measured at the midpoint of the stressed vowels in *fought* and *God* (both in the final paragraph of the "Gettysburg Address").⁵ If the difference was less than 0.5, indicating that the vowels in the two words were relatively close in height, the speaker was excluded. The speakers who met both criteria (n = 45 / 172, or 26.2%) were included in our analysis.

The following vowel plots illustrate the effect of our selection criteria. Note that they include a subset of /æ/ tokens, as well as the words *cause*, *fought*, *God*, and *fathers*, which contain the relevant vowels that figured into the criteria for New York City. The plots do not reflect the full vowel space; for example, the peripheral vowels /i/ and /u/ are excluded. Note also that formant measurements for the /æ/ tokens were taken at the point in time where the maximum F2 was reached, while the other tokens were measured at their midpoints. Finally, the plots show the average value for F1 and F2 within each state based on speakers' raw Bark values; the statistical analysis and discussion of vowel trajectories presented later in this paper are based on normalized values.

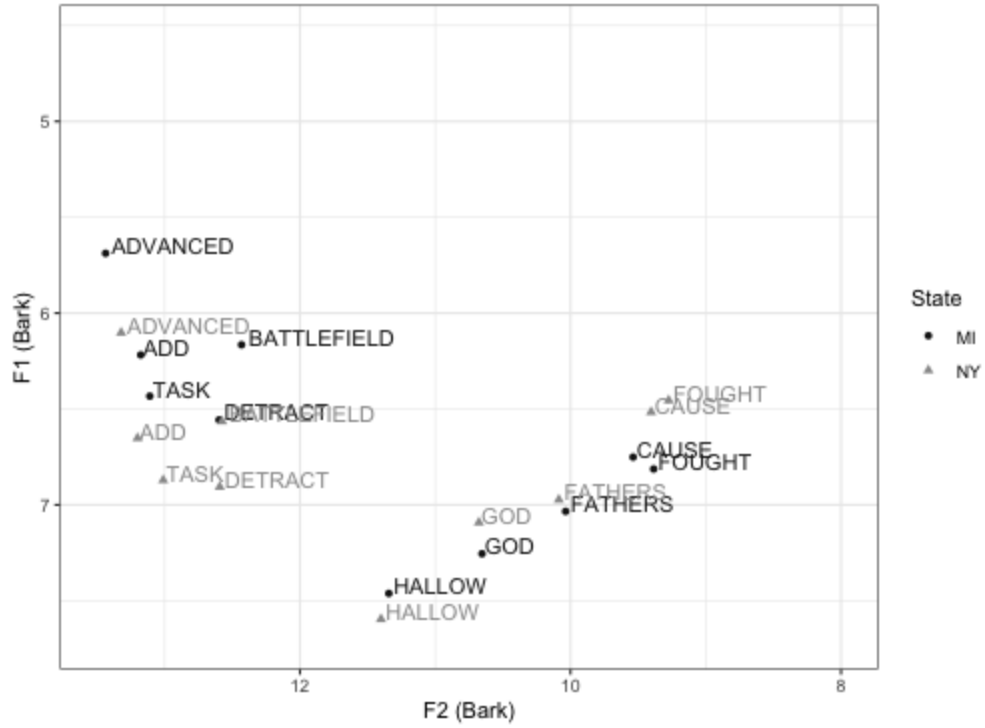


Figure 1. Partial vowel plot for all 402 speakers from MI and NY

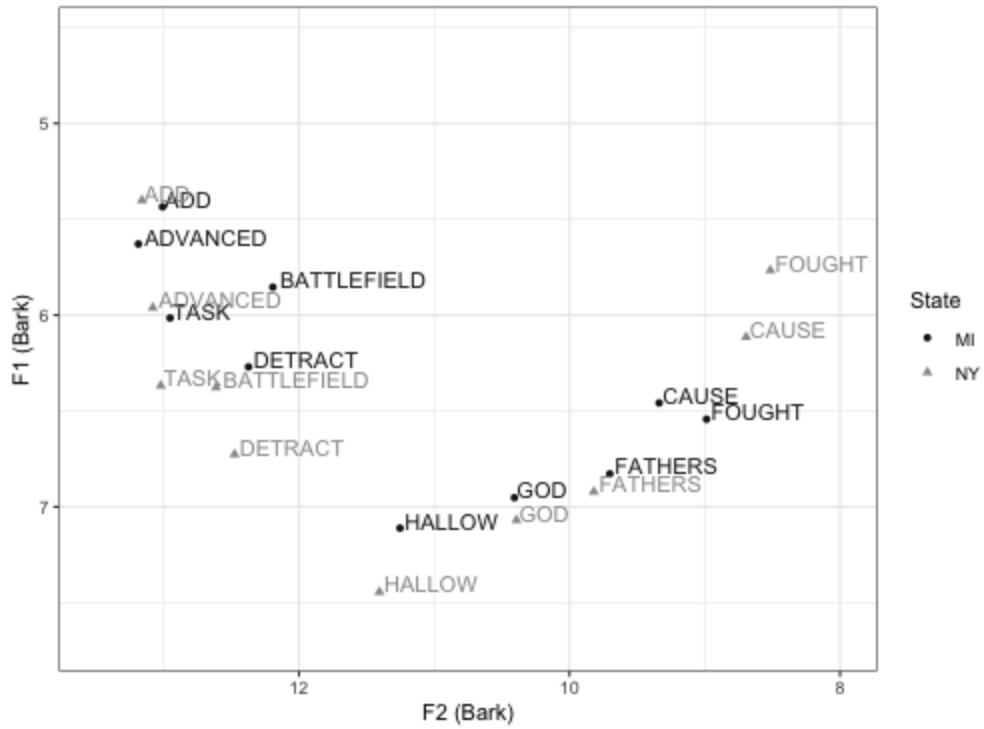


Figure 2. Partial vowel plot for the 157 speakers from MI and NY included in the study

As the plots show, the speakers who satisfied both selection criteria for inclusion in our study (157 / 402 or 39.1%) exhibit a general spreading out of the vowel space. We see that speakers included from Michigan have comparatively raised and fronted vowels for both *add* and *advanced*; this is to be expected, of course, because we specifically relied on the difference between these items as a selection criterion. Note, though, that all of the other items with stressed /æ/ are also higher for the included speakers, which is consistent with the goal of our selection procedure to target dialect speakers (e.g., *task* is now much closer to *advanced*). For the New Yorkers who satisfied the selection criteria, we find a more dramatic difference between *fought* (/ɔ/) and *God* (/ɑ/) than expected; this difference also generalizes to *cause* and *fathers*, words that did not play a role in the selection criteria. Finally, the location of *hallow* in the vowel space presumably reflects the fact that many speakers in both state groups pronounced the word as a homophone of *hollow*, i.e., with stressed /ɑ/ rather than /æ/.

In the discussion that follows, we use the terms “Michiganders” and “New Yorkers” to refer to the speakers who satisfied our criteria for inclusion in the study. Because these criteria were designed to capture traditional acoustic properties of the Inland North and New York City dialects, we take these speakers to be representative of those dialects. However, these categories were obtained in an automated fashion, with no background information about the speakers other than their state of residence. The categories should therefore be interpreted with caution and should not be viewed as a definitive claim that all speakers included from the state of Michigan are Inland North speakers, or that all speakers included from the state of New York are New York City dialect speakers.

3.3. Usability of Forced Aligned Data

Before turning to the study of the phonetic realizations of /æ/-tensing by Michiganders and New Yorkers, it is important to establish the usability of our forced aligned data. Because of the inherent disadvantages of YouTube recordings discussed above—the lack of demographic information on speakers, the variation in recording equipment, the quality of audio compression, etc.—any acoustic data extracted from YouTube recordings can never be considered fully *reliable*, in the sense of being commensurate in quality to data gathered from sociolinguistic interviews. What we can do, however, is establish that the data are *usable*, in the sense that they can be used to generate hypotheses that can subsequently be tested with more established methods and data sources. The use of so-called “bad data” has been advocated before in sociolinguistics, especially to address historical questions for which new data cannot be collected (Nevalainen 1999; Hickey 2017). For questions of contemporary dialect differences, for which higher quality data *can* be collected, the use of social media recordings is a less resource-intensive way to generate hypotheses that can guide additional research—even if the data are not the most high-quality in absolute terms. The focus in this section is to establish the usability of the *Gettysburg Corpus* data.

In all /æ/-tensing systems in American English, the vowel has the lowest F1 frequency (an acoustic correlate of increased vowel height) and highest F2 frequency (an acoustic correlate of increased frontness) when followed by a nasal (e.g., Labov, Ash, and Boberg 2006, 176; Dinkin 2011). Visual inspection of the vowel space of the initial corpus prior to speaker selection (Figure 1) indicates that the F1 of /æ/ in the word *advanced* is indeed lower than in other

contexts (i.e., it is a higher vowel), while F2 is greater (i.e., a fronter vowel). The position of *advanced* in the vowel space is largely maintained after speaker selection (Figure 2). Our automatically compiled and forced aligned corpus thus replicates one finding from the dialectological literature on /æ/-tensing.

As a second check on the usability of the forced alignment of the *Gettysburg Corpus*, we also replicated the finding that vowels are longer when preceding a voiced obstruent than when preceding a voiceless one (e.g., Chen 1970). Using the original data set, prior to the selection of participants described above, we extracted all word tokens that included a single stressed vowel (not just /æ/) preceding a simple obstruent coda. This was achieved by producing a subset of the data containing only monosyllabic words for which FAVE's orthographic transcription of the vowel ended in a "1" for primary stress (thus avoiding polysyllabic words with secondary stress as well as unstressed function words), and then subsetting the data further to include only those tokens with simple codas that ended in an obstruent. This resulted in 16,482 stressed monosyllables with obstruent codas (402 speakers, 41 tokens each). Because there are inherent differences in duration across vowel phonemes (e.g., high vowels tend to be shorter than low vowels; Peterson and Lehiste 1960) and across individuals (e.g., based on speech rate), a linear mixed-effects regression model was produced treating following environment (voiced vs. voiceless obstruent) and state (MI vs. NY) as fixed factors, and vowel and speaker as random factors. (Word context was excluded as a factor because it is largely collinear with vowel and following environment, due to the relatively small number of unique words in the "Gettysburg Address.") As expected, the model finds that duration increases significantly in vowels preceding a voiced obstruent over those preceding a voiceless obstruent. There is no significant effect of

state, either as a main predictor or in interaction with following environment (tested in a separate model not shown here). The model summary is provided in Table 1, and the model’s predictions for following environment and state are also visualized in Figure 3.

	<u>Estimate (β)</u>	<u>Std. error</u>	<u>t-value</u>	<u>p-value</u>	<u>N</u>
Intercept	155.47	17.95	8.66	< 0.01	16,482
Following environment (vs. voiced obstruent)					
voiceless obstruent	-49.71	1.64	-30.28	< 0.01	8,844 7,638
State (vs. MI)					
NY	5.06	3.04	1.67	0.10	9,430 7,052

Table 1. Summary of linear mixed-effects model for all stressed monosyllables with obstruent codas

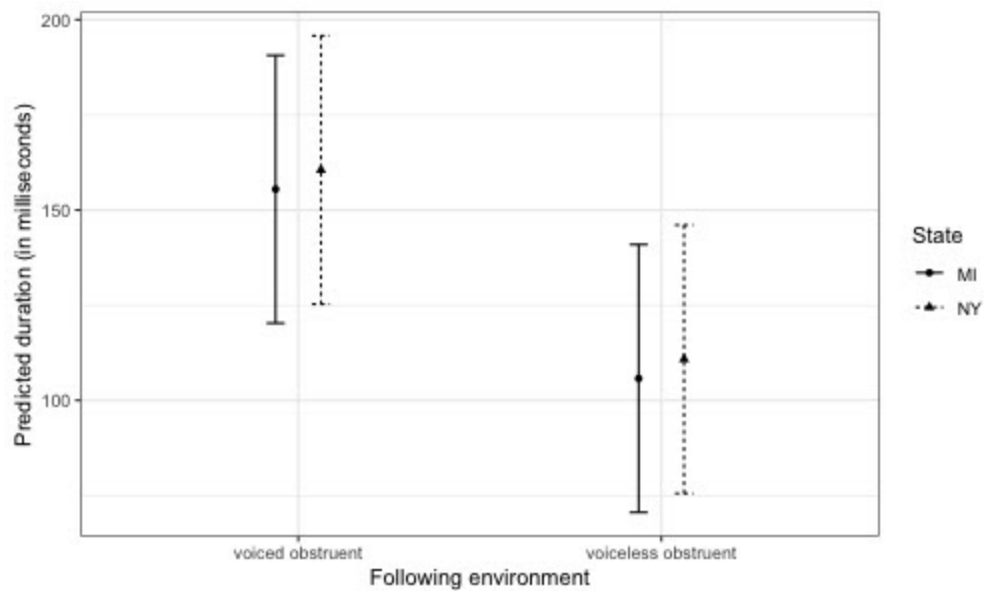


Figure 3. Model-predicted duration for vowels in monosyllables by following environment (significant diff.) and state (not significant diff.)

4. Formant Trajectory of /æ/

The fact that the methodology outlined above is capable of replicating some broad findings of previous studies inspires confidence that the *Gettysburg Corpus* can make a real contribution to novel research questions, as well. In this section we use the recitation data to examine the regional and linguistic factors influencing the formant trajectory of /æ/.

4.1. Methods

As described previously, the current study relied on automatically extracted measurements of /æ/ tokens from the recitations by Michiganders and New Yorkers (those speakers who satisfied our selection criteria). F1, F2, and F3 (in Bark) had been extracted at 10% intervals over the duration of the vowel for forced aligned /æ/ tokens (from 10% to 90%) and normalized using the Bark Difference Metric (Syrdal and Gopal 1986; Thomas and Kendall 2007). Of interest were the lexical items *add*, *advanced*, *last*, and *task*, which represent three environments (before a voiced stop, nasal, and voiceless fricative coda) said to be tense in both the Inland North and New York City dialects. Note that the relevant vowel in each lexical item precedes a coronal consonant; only the manner of articulation, then, varies across the tokens. In total, 112 tokens from Michigan and 45 tokens from New York (i.e., one per recitation) were obtained for each lexical item (n = 627, with 9 formant measurements per token).⁶

Tokens were measured at multiple points to allow us to consider how the vowel changed

across the formant trajectory. It is well known that acoustic properties outside of the spectral characteristics of a single point contribute to distinguishing a vowel both in production and perception (duration: Clopper, Pisoni, and de Jong 2005; Wassink 2006; Fridland, Kendall, and Farrington 2014; Vowel Inherent Spectral Change [VISC]: Fox and Jacewicz 2009; Van der Harst, Van de Velde, and Van Hout 2014; Farrington, Kendall, and Fridland 2018). This is true even of monophthongs (Morrison and Assmann 2013), and so it is particularly important to consider properties like VISC in the case of /æ/. There are several possible methods for quantifying VISC in data analysis. Fox and Jacewicz (2009) advocate collapsing spectral measurements of a few points in the vowel duration into measures of vector length, trajectory length, and the spectral rate of change, while Van der Harst, Van de Velde, and Van Hout (2014) test the possibility (among others) of using regression analysis to obtain one overall measure.

Other studies fit a curve to multi-point spectral measurements to obtain a more holistic result, which is the approach we pursue here. We applied smoothing spline analysis of variance (SS ANOVA) to analyze vowel trajectories. SS ANOVA is used to estimate whether there is a significant difference between curves. Davidson (2006) introduces its use for linguistics with regard to tongue shape measurements in ultrasound studies, and recent studies have extended its use to an analysis of formant trajectories across vowels (Nycz and De Decker 2006; Docherty, Gonzalez, and Mitchell 2015; Hall 2016). The procedure, as implemented in the *gss* package in R (Gu 2014), plots a smoothed cubic fit to the data and constructs Bayesian confidence intervals around the curve. This enables the researcher to determine not only whether curves differ from one another, but when in time those differences occur. While Morrison (2013) contends that such curves perform no better at quantifying VISC than measurements of two points, he also

acknowledges that using fewer measurements may risk ignoring information pertinent to the perception of the vowel, including socially indexical information (Morrison 2013, 45). We elected to use SS ANOVA to emphasize our focus on whether the formant curves differ at any point in time during vowel production, in order to minimize the risk of losing such information. We additionally note that SS ANOVA inherently normalizes our data for duration. We do not believe this will cause issues for our analysis; as found above, there is no significant difference between Michiganders and New Yorkers in vowel duration.

We use SS ANOVA to compare formant trajectories for tense /æ/ between states and between contexts. Because confidence intervals are constructed around the curve, the graph of the model serves as not only a visualization of the data, but a result as well, with areas of non-overlap representing time points with significant differences. We present such graphs along with the r^2 value, in order to show where and how the groups differ, as well as how much variation is captured by the model. Separate models are run for F1 and F2, with following environment, speaker's state residency, and the measurement point in time as independent variables. We consider models in which interactions between factors are considered, as well as models without interactions. For both formants, the model with the lowest mean squared error, in which all interactions are included, was chosen.

4.2. Results

SS ANOVA shows main effects for state and context, in addition to a significant interaction of state, context, and point in the vowel's duration ($r^2 = 0.186$ for F1, $r^2 = 0.204$ for F2). These effects are illustrated in Figures 4 and 5, with Bark difference normalized F1

displayed in the bottom half of the figure and F2 in the top half. Lower normalized values correspond to lower and fronter vowels, while higher normalized values correspond to higher and backer vowels. Because we reverse the y-axis in these figures, interpreting them is not unlike reading formant trajectories from a spectrogram. In the following descriptions, we follow the convention of discussing changes in F1/F2 in terms of “lowering” and “backing” with the caveat that we do not have any articulatory data for our speakers. As shown in Figure 4, there is a significant difference between state groups in F1 of /æ/ across the entirety of the vowel duration, as Michiganders have a lower value (indicating a lower vowel) at all points. Both groups lower the vowel over time; however, Michiganders appear to lower it more than New Yorkers, such that the difference between groups at 75% of the vowel duration is larger than that at 25% of the vowel duration. The Michiganders begin F2 of /æ/ with a lower value (a fronter vowel) than the New Yorkers and back over time to end with a value of F2 indistinguishable from the New Yorkers. While the New Yorkers also back the vowel over time, their F2 trajectory is flatter than the Michiganders’. The trajectory of both F1 and F2, particularly among the Michiganders, is consistent with the diphthongization of Northern breaking ([ɛə] or [ɪə]), as the lowering and backing effect could be attributable to tongue retraction over the course of production. The main effect thus indicates regional differences in the inglide, which could be due to either a difference in the magnitude of the inglide (Michiganders retract the vowel lower and further back than do New Yorkers) or to variation in how consistently an inglide is found across speakers and contexts (Michiganders produce a diphthong more often than do New Yorkers).

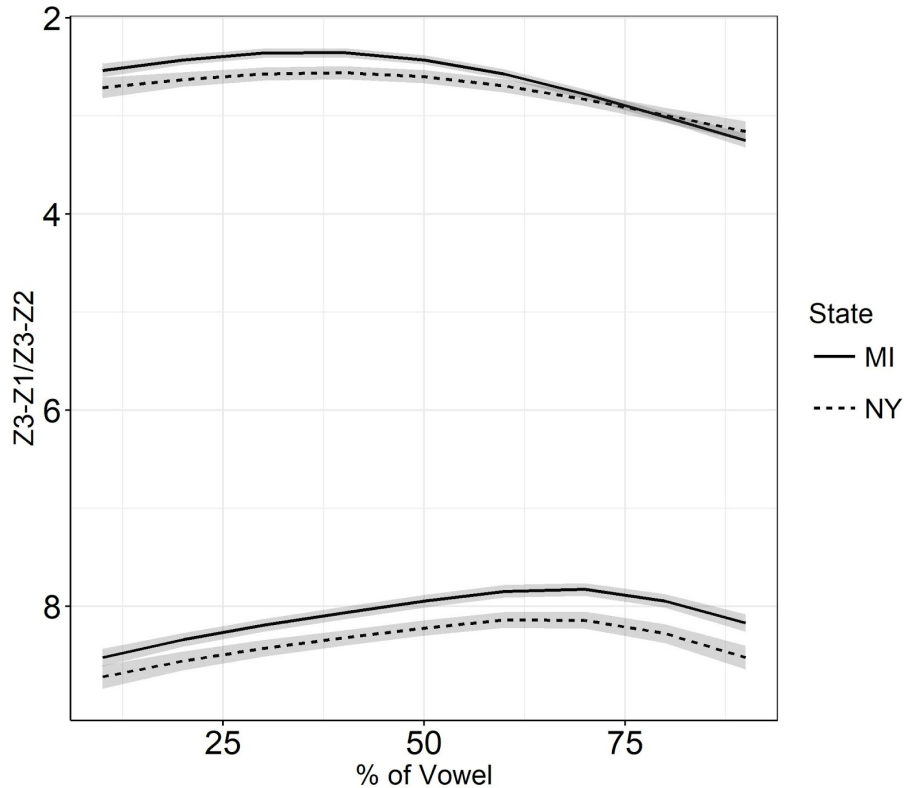


Figure 4. SS ANOVA comparing formant trajectory of /æ/ in two states, aggregated over all contexts

The state-context interaction is illustrated in Figure 5. The results here show both a main effect of context and that the difference between state groups varies from context to context. Unsurprisingly, SS ANOVA indicates that /æ/ has a higher F1 and lower F2 value (indicating a higher and fronter vowel) in *advanced* than in the other contexts. This is particularly evident in the first half of the vowel. The main effect of context also involves the formant trajectory. /æ/ exhibits the most movement in *add* and *advanced*, while the vowel is essentially a monophthong in *last*. Given the formant trajectory of *task*, this latter effect may be due in part to coarticulation with preceding /l/. In *advanced*, the vowel backs dramatically and lowers slightly across the vowel duration. By contrast, the vowel lowers more in *add*, and this lowering begins earlier in

the vowel duration than in *advanced*. This main effect indicates that tense /æ/ is not realized uniformly in all tensing contexts.

Turning to the significant state-context interaction, we first see that F1 begins as more or less indistinguishable between groups, but the groups quickly separate. While the Michiganders have a lower /æ/ in the second half of the vowel in all contexts, the differences between groups is much greater in *add* than in other contexts. The state effect for F1 thus varies in degree by context. In F2, whether the state effect is found varies categorically by context. It is absent in *add* and *last*, as the two groups have roughly identical formant trajectories. In contrast, the Michiganders begin the vowel significantly fronter than the New Yorkers in *advanced* and *task*, before backing the vowel to match the New Yorkers over the course of the vowel duration. This effect is most pronounced in *advanced*. In sum, although the New Yorkers do show an inglide (F2 raises in value and F1 lowers in value over the course of the vowel, suggesting potential tongue lowering/retraction), the movement is much more pronounced among Michiganders in a way that is consistent with the Northern breaking-style diphthong. The Michiganders vary, however, in how the movement occurs by context. In *add*, the retraction is clearest in F1, while in *advanced*, it is most evident in F2. The formant trajectories of both contexts are consistent with a diphthongal production.

While we find significant differences in production, we do note the overall similarity between groups. This of course can be attributed in part to our selection criteria. By setting a criterion based on a static measure of F1 of *add* and *advanced*, we would expect the groups to be quite similar to each other with respect to this value at a comparable point in time. However, the fact that the F1 similarity extends to *last* and *task*, that the similarity of both formants persists

across the vowel duration, and that F2 resolves quite similarly in each context appears to be too systematic to be attributable solely to our selection method, as we made no decisions regarding speaker selection based on any of these outcomes. Note too that the groups appear to have quite similar F1 *and* F2 at the peak of F2 for each context. A single point measurement of the vowel would yield the results that there is little difference between groups; the difference only becomes quite apparent by collecting measurements at multiple time points.

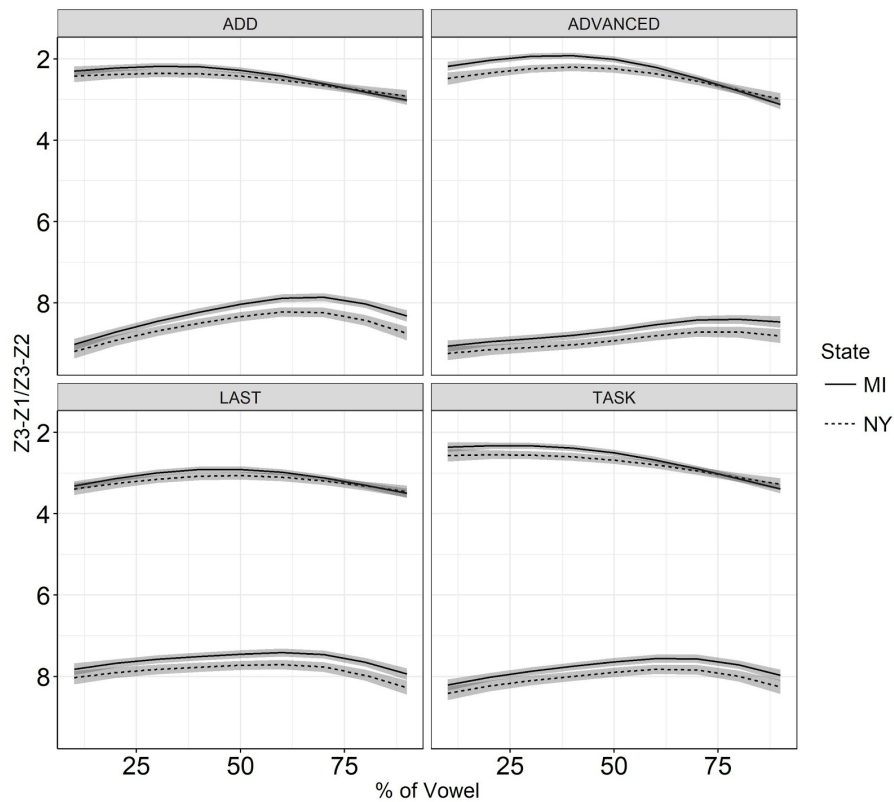


Figure 5. SS ANOVA showing state-context interaction in formant trajectory of /æ/

5. Discussion

The results of our corpus study suggest that what has been characterized as “tense /æ/” (or labeled as “/æh/”) does in fact differ across dialects and contexts. Both New Yorkers and Michiganders display what Labov, Ash, and Boberg (2006) would characterize as inglides (that

is, both groups lower and back the vowel over the course of its duration), but Michiganders have significantly larger inglides in both F1 and F2 than do New Yorkers. Similarly, the formant trajectory varies across contexts, as Michiganders retract more in F2 in *advanced* but more in F1 in *add*. These findings in particular call into question whether “/æ/-tensing” is a coherent cross-dialectal phenomenon—an assumption implicit in much of the literature on dialectal variation in vowel systems but never before interrogated using online corpus data.

The movement in F1 and F2 within the group of Michiganders (and New Yorkers, though to a much lesser extent) is potentially consistent with the Northern breaking pattern. However, it is important to note that this movement is realized differently in different contexts. In addition to the aforementioned effects, there is less movement when /æ/ precedes a voiceless fricative, suggesting that Northern breaking may not be found as often in such an environment. Despite these findings, there is no clear evidence indicating a second mora in the vowel in the sense of the vowel being composed of two short, steady-state vowels of equal length (cf. Labov, Ash, and Boberg 2006, 177). This can be attributed to two factors: individual variation and the nature of smoothing splines. First, when it has been observed, Northern breaking appears to vary alongside the late ingliding variant, even in the speech of an individual speaker (Labov, Ash, and Boberg’s [2006] example of Sharon K. of Rochester, mentioned above). The use of both Northern breaking and late ingliding could have the effect of erasing clear evidence of a diphthong with distinct steady states. Second, smoothing splines by nature are designed to optimize fit to the data and smoothness. Given the rather low number of measurements per token used for the SS ANOVA in this study (i.e., one measurement at each 10% interval; by contrast, Davidson [2006] uses 64 points per token, while Nycz and De Decker [2006] use 50 per token), it is possible that

sharper inglides between measurements were smoothed out. To avoid this issue, further corpus studies might consider a larger number of measurements per vowel token.

That said, there is clearly a difference between tense /æ/ across the Michiganders and New Yorkers. The New Yorkers do display some evidence of backing and lowering across the vowel duration, but not to the same degree as Michiganders. Put another way, the New Yorkers display more consistency in formant trajectory across environments than do the Michiganders. This consistency appears to suggest that New Yorkers engage in less Northern breaking (if any at all) than do Michiganders. Our analysis does not show, however, whether this is a difference in degree or in rate. In any case, we suggest that Northern breaking is primarily limited to the Inland North regionally and to pre-nasal and pre-voiced stop contexts environmentally.

One generalization frequently made about American English is that all dialects exhibit /æ/-tensing in pre-nasal contexts. Indeed, we find that both groups have raised and fronted tokens in this context which lower and back across the duration of the vowel. At the same time, there is a regional difference in the implementation of pre-nasal /æ/-tensing: during the inglide, Michiganders back and lower the vowel more than New Yorkers do, as evidenced by the significant differences in F2 in the first half of the vowel and in F1 in the second half of the vowel. The fact that we find such a regional difference in the implementation of tensing even in the same pre-nasal context is intriguing. If our study of Inland North and New York City speakers demonstrates that the tense variant is not uniform across dialects, what does it look like in dialects with nasal systems? We might hypothesize that if the Northern breaking variant is regionally restricted, then the variant that would be found in nasal systems elsewhere will more closely approximate that found among New Yorkers. However, because the general

implementation appears to be largely shared, the degree of retraction may show yet more regional variation when other short-*a* systems are considered. This is a question worth exploring further as it has implications for phonological theory, since phonologists who discuss /æ/ often treat all tense variants as identically produced diphthongs (see, e.g., Green 2001; Duncan 2016; Nie 2017). If this is not the case even in pre-nasal contexts, such descriptions may be oversimplified and in need of revision.

6. Conclusion

An important question that our computer-automated methodology raises is how reliable the results are, particularly because the data are so noisy. We find that the *Gettysburg Corpus* replicates dialectological work showing that /æ/ has its highest F1 value and lowest F2 value (corresponding to being highest/frontest in production) in pre-nasal contexts, as well as phonetic work showing that vowels are longest when preceding voiced obstruents. This gives us confidence that the context- and region-based differences in formant trajectory identified in our study can be attributed to actual differences, rather than to errors introduced by our automation methods.

Given these preliminary replications, we argue that it is possible to utilize large online speech corpora (e.g., YouTube videos) for acoustic analysis, even if such corpora were not compiled for that purpose and, perhaps more importantly, despite the great deal of noise in the data. The results thus open a path for further research not only on the various articulations of /æ/, but also on other sociophonetic topics that are well-suited for so-called “big data” corpus analysis. These computational methods are especially appropriate for addressing novel research

problems that require large numbers of speakers. Social media sites are promising as sources of sociophonetic data, especially in cases where transcripts are readily available (e.g., TED talks) and can be fed into forced alignment software. Their use, however, raises important questions for sociolinguistic theory. Recall, for example, that we applied post-hoc criteria to select an appropriate sample of dialect speakers from the pool of residents of Michigan and New York State. What are the effects of these criteria on our final sample? Did we, in fact, accurately screen for dialect region? What, if any, are the social correlates (e.g., gender, age, race and ethnicity) of such numerical criteria? These are all questions worth exploring in a careful comparative study. In the interim, we advocate the use of corpus methods to refine existing hypotheses about language variation, generate new ones, and supplement—rather than replace—sociophonetic data gathered through more established methods, including sociolinguistic interviews and controlled reading tasks.

Notes

An earlier version of this paper was presented at NWAV 45. Thank you to three anonymous reviewers, Lisa Davidson, Laurel MacKenzie, John V. Singler, and graduate students in the Acoustic Phonetics course at New York University for thoughts and comments, as well as Zack Jagers for advice on glide measurements. This project would not have been possible without the existence of LearnTheAddress.org.

1. The audio extractor script, *youtube-dl*, can be found here:

<https://rg3.github.io/youtube-dl/>. Approximately thirty videos had broken links (e.g., the

videos had been deleted or set to private) and could not be used. A handful were of a particular shortened URL format (i.e., “youtu.be”) and were also not included.

2. The procedure described here was carried out on March 3, 2016, and reflects the number of videos that were cataloged at LearnTheAddress.org at that time.
3. An anonymous reviewer for *American Speech* expressed concern about the possibility that our automation methods produced an imbalance in speaker sex between the sample from Michigan and the sample from New York. While our goal is to test the feasibility of a maximally automated approach to data collection, at the reviewer’s suggestion we hand-checked all YouTube recordings included in our final data set (157 in total) and tabulated speaker sex, based on their gender presentation. There was no statistically significant difference between Michigan and New York (about three-quarters of the speakers from each state were men). Furthermore, only two speakers appeared to have memorized the “Gettysburg Address,” while all the rest were reading the text—none in a manner that we would characterize as particularly performative. Three speakers wore top hats, but none of these speakers had memorized the “Address.”
4. “The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to *add* or detract... It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly *advanced*.”
5. “It is for us the living, rather, to be dedicated here to the unfinished work which they who *fought* here have thus far so nobly advanced... [We here highly resolve] that this nation, under *God*, shall have a new birth of freedom.”
6. One token of *last* is missing from a single New York speaker, yielding 44 tokens for this

context from the NY group.

References

- Becker, Kara. 2010. "Regional Dialect Features on the Lower East Side of New York City: Sociophonetics, Ethnicity, and Identity." Ph.D. dissertation, New York University.
<https://pqdtopen.proquest.com/pubnum/3428032.html>.
- Becker, Kara. 2016. "Linking Community Coherence, Individual Coherence, and Bricolage: The Co-Occurrence of (r), Raised BOUGHT and Raised BAD in New York City English." *Lingua* 172–3: 87–99. <https://doi.org/10.1016/j.lingua.2015.10.017>.
- Becker, Kara, and Amy Wing-mei Wong. 2010. "The Short-*a* System of New York City English: An Update." *University of Pennsylvania Working Papers in Linguistics* 15.2: 11–20.
<http://repository.upenn.edu/pwpl/vol15/iss2/3>.
- Boersma, Paul, and David Weenink. 2017. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.0.28. <http://www.praat.org>.
- Bulgin, James, Paul De Decker, and Jennifer Nycz. 2010. "Reliability of Formant Measurements from Lossy Compressed Audio." Poster presentation at the *British Association of Academic Phoneticians Colloquium*, London. <http://research.library.mun.ca/id/eprint/684>.
- Callary, Robert E. 1975. "Phonological Change and the Development of an Urban Dialect in Illinois." *Language in Society* 4.2: 155–69. <https://doi.org/10.1017/S0047404500004620>.
- Cassidy, Frederic G., ed. 1985. *Dictionary of American Regional English*, vol. 1. Cambridge, MA: Belknap/Harvard University Press.
- Chen, Matthew. 1970. "Vowel Length Variation as a Function of the Voicing of the Consonant Environment." *Phonetica* 22.3: 129–59. <https://doi.org/10.1159/000259312>.

- Clopper, Cynthia G. 2009. "Computational Methods for Normalizing Acoustic Vowel Data for Talker Differences." *Language and Linguistics Compass* 3.6: 1430–42.
<https://doi.org/10.1111/j.1749-818X.2009.00165.x>.
- Clopper, Cynthia G., David B. Pisoni, and Kenneth de Jong. 2005. "Acoustic Characteristics of the Vowel Systems of Six Regional Varieties of American English." *Journal of the Acoustical Society of America* 118.3: 1661–76. <https://doi.org/10.1121/1.2000774>.
- Coggshall, Elizabeth Learn. 2017. "Short-*a* in the Sixth Borough: A Sociophonetic Analysis of a Complex Phonological System in Jersey City." Ph.D. dissertation, New York University.
<https://search.proquest.com/docview/1941439935>.
- Cohen, Paul Sheldon. 1970. "The Tensing and Raising of 'Short A' in the Metropolitan Area of New York City." Master's thesis, Columbia University.
- D'Onofrio, Annette, and Jaime Benheim. 2019. "Contextualizing Reversal: Local Dynamics of the Northern Cities Shift in a Chicago Community." *Journal of Sociolinguistics*. Advance online publication. <https://doi.org/10.1111/josl.12398>.
- Davidson, Lisa. 2006. "Comparing Tongue Shapes from Ultrasound Imaging Using Smoothing Spline Analysis of Variance." *Journal of the Acoustical Society of America* 120.1: 407–15. <https://doi.org/10.1121/1.2205133>.
- De Decker, Paul, and Jennifer Nycz. 2011. "For the Record: Which Digital Media Can Be Used for Sociophonetic Analysis?" *University of Pennsylvania Working Papers in Linguistics* 17.2: 51–9. <http://repository.upenn.edu/pwpl/vol17/iss2/7>.
- De Decker, Paul M., and Jennifer R. Nycz. 2012. "Are Tense [æ]s Really Tense? The Mapping Between Articulation and Acoustics." *Lingua* 122.7: 810–21.

- <https://doi.org/10.1016/j.lingua.2012.01.003>.
- Dinkin, Aaron J. 2011. "Nasal Short-*a* Systems vs. the Northern Cities Shift." *University of Pennsylvania Working Papers in Linguistics* 17.2: 71–80.
<http://repository.upenn.edu/pwpl/vol17/iss2/9>.
- Dinkin, Aaron J. 2013. "Settlement Patterns and the Eastern Boundary of the Northern Cities Shift." *Journal of Linguistic Geography* 1.1: 4–30. <https://doi.org/10.1017/jlg.2013.2>.
- Docherty, Gerard, Simón Gonzalez, and Nathaniel Mitchell. 2015. "Static vs. Dynamic Perspectives on the Realization of Vowel Nucleii [sic.] in West Australian English." *Proceedings of the 18th International Congress of Phonetic Sciences*.
<https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0956.pdf>.
- Driscoll, Anna, and Emma Lape. 2015. "Reversal of the Northern Cities Shift in Syracuse, New York." *University of Pennsylvania Working Papers in Linguistics* 21.2: 41–7.
<http://repository.upenn.edu/pwpl/vol21/iss2/6/>.
- Duncan, Daniel. 2016. "'Tense' /æ/ is Still Lax: A Phonotactics Study." *Proceedings of the 2015 Annual Meeting on Phonology*: 1–12. <https://doi.org/10.3765/amp.v3i0.3653>.
- Eisenstein, Jacob, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2014. "Diffusion of Lexical Change in Social Media." *PLoS ONE* 9.11: e113114.
<https://doi.org/10.1371/journal.pone.0113114>.
- Farrington, Charlie, Tyler Kendall, and Valerie Fridland. 2018. "Vowel Dynamics in the Southern Vowel Shift." *American Speech* 93.2: 186–222.
<https://doi.org/10.1215/00031283-6926157>.

- Fox, Robert Allen, and Ewa Jacewicz. 2009. "Cross-Dialectal Variation in Formant Dynamics of American English Vowels." *Journal of the Acoustical Society of America* 126.5: 2603–18. <https://doi.org/10.1121/1.3212921>.
- Fridland, Valerie, Tyler Kendall, and Charlie Farrington. 2014. "Durational and Spectral Differences in American English Vowels: Dialect Variation Within and Across Regions." *Journal of the Acoustical Society of America* 136.1: 341–9. <https://doi.org/10.1121/1.4883599>.
- Goodheart, Jill C. 2004. "I'm No *Hoosier*! Evidence of the Northern Cities Shift in St. Louis, Missouri." Master's thesis, Michigan State University. <https://doi.org/10.25335/M58K7555S>.
- Green, Antony Dubach. 2001. "The Tense-Lax Distinction in English Vowels and the Role of Parochial and Analogical Constraints." *Linguistics in Potsdam* 16: 32–57. <https://doi.org/10.7282/T3CJ8G70>.
- Gu, Chong. 2014. "Smoothing Spline ANOVA Models: R Package *gss*." *Journal of Statistical Software* 58.5: 1–25. <http://doi.org/10.18637/jss.v058.i05>.
- Guy, Gregory R. 2013. "The Cognitive Coherence of Sociolects: How Do Speakers Handle Multiple Sociolinguistic Variables?" *Journal of Pragmatics* 52: 63–71. <https://doi.org/10.1016/j.pragma.2012.12.019>.
- Haddican, William, Christina Tortora, Michael Newman, Cecelia Cutler, C. E. A. Diertani, Leyla Eldridge, Sabrina Lagreca, and Zi Zi Gina Tan. 2018. "Systemic Change and Parent First-Dialect Effects in NYC English Short-*a* Variation." Paper presentation at *New Ways of Analyzing Variation* 47, New York.

- Hall, Erin. 2016. "Static and Dynamic Analyses of Canadian Raising in Toronto and Vancouver." Paper presentation at *New Ways of Analyzing Variation* 45, Vancouver.
- Hickey, Raymond. 2017. "Analysing Early Audio Recordings." In Raymond Hickey (ed.), *Listening to the Past: Audio Records of Accents of English*, 1–12. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107279865.002>.
- Hooper, Joan Bybee. 1977. "Substantive Evidence for Linearity: Vowel Length and Nasality in English." *Papers from the Regional Meeting of the Chicago Linguistic Society* 13: 152–64.
- Huang, Yuan, Diansheng Guo, Alice Kasakoff, and Jack Grieve. 2016. "Understanding U.S. Regional Linguistic Variation with Twitter Data Analysis." *Computers, Environment and Urban Systems* 59: 244–55. <https://doi.org/10.1016/j.compenvurbsys.2015.12.003>.
- Jacewicz, Ewa, and Robert Allen Fox. 2013. "Cross-Dialectal Differences in Dynamic Formant Patterns in American English Vowels." In Geoffrey Stewart Morrison and Peter F. Assmann (eds.), *Vowel Inherent Spectral Change*, 177–98. Berlin: Springer. https://doi.org/10.1007/978-3-642-14209-3_8.
- Kurath, Hans. 1949. *A Word Geography of the Eastern United States*. Ann Arbor: University of Michigan Press.
- Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov, William. 1989. "Exact Description of the Speech Community: Short *a* in Philadelphia." In Ralph W. Fasold and Deborah Schiffrin (eds.), *Language Change and Variation*, 1–57. Amsterdam/Philadelphia: John Benjamins.

- Labov, William. 1994. *Principles of Linguistic Change*, vol. 1: *Internal Factors*. Malden, MA: Blackwell.
- Labov, William. 2007. "Transmission and Diffusion." *Language* 83.2: 344–87.
<https://doi.org/10.1353/lan.2007.0082>.
- Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.
- Labov, William, Sabriya Fisher, Duna Gylfadottir, Anita Henderson, and Betsy Sneller. 2016. "Competing Systems in Philadelphia Phonology." *Language Variation and Change* 28.3: 273–305. <https://doi.org/10.1017/S0954394516000132>.
- Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. "One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis." *Language* 89.1: 30–65. <https://doi.org/10.1353/lan.2013.0015>.
- Milholland, Agatha. 2018. "Reversal of the Northern Cities Shift in Buffalo, NY." Paper presentation at *New Ways of Analyzing Variation* 47, New York.
- Morrison, Geoffrey Stewart. 2013. "Theories of Vowel Inherent Spectral Change." In Geoffrey Stewart Morrison and Peter F. Assmann (eds.), *Vowel Inherent Spectral Change*, 31–47. Berlin: Springer. https://doi.org/10.1007/978-3-642-14209-3_3.
- Morrison, Geoffrey Stewart, and Peter F. Assmann, eds. 2013. *Vowel Inherent Spectral Change*. Berlin: Springer. <https://doi.org/10.1007/978-3-642-14209-3>.
- Nesbitt, Monica. 2018. "Changing Accents Changes Syllables: The Effect of Diachronic Phonetic Change on Syllabic Representations in the Inland North." Paper presentation at *New Ways of Analyzing Variation* 47, New York.

- Nevalainen, Terttu. 1999. "Making the Best Use of 'Bad' Data: Evidence for Sociolinguistic Variation in Early Modern English." *Neuphilologische Mitteilungen* 100.4: 499–533.
<https://www.jstor.org/stable/43346227>.
- Newman, Michael. 2014. *New York City English*. Berlin: De Gruyter Mouton.
- Nie, Yining. 2017. "Phonetic Enhancement and Three Patterns of English *a*-Tensing." *University of Pennsylvania Working Papers in Linguistics* 23.1: 181–90.
<http://repository.upenn.edu/pwpl/vol23/iss1/21>.
- Nycz, Jennifer, and Paul De Decker. 2006. "A New Way of Analyzing Vowels: Comparing Formant Contours Using Smoothing Spline ANOVA." Poster presentation at *New Ways of Analyzing Variation* 35, Columbus, OH. <http://research.library.mun.ca/655/>.
- Peterson, Gordon E., and Ilse Lehiste. 1960. "Duration of Syllable Nuclei in English." *Journal of the Acoustical Society of America* 32.6: 693–703. <https://doi.org/10.1121/1.1908183>.
- R Core Team. 2017. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rankinen, Wil. 2018. "Lingering Substrate and Encroaching Exogenous Influences on Finnish and Italian Americans' Vowels in Michigan's Marquette County." *American Speech* 93.2: 223–69. <https://doi.org/10.1215/00031283-7115889>.
- Rankinen, Wil, Aaron Albin, and T. J. Neuhaus. 2019. "Decline of the Northern Cities Vowel Shift in Western Lower Michigan: Apparent-Time Evidence of a Change in Progress." Paper presentation at the *American Dialect Society*, Salt Lake City, UT.
- Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2014. "FAVE (Forced Alignment and Vowel Extraction)

- Program Suite,” v1.2.2. <https://doi.org/10.5281/zenodo.22281>.
- Shapp, Allison. 2019. “A Sociophonetic Analysis of New York City English in the Speech of Teenagers from Nassau County, Long Island.” Ph.D. dissertation, New York University.
- Syrdal, Ann K., and H. S. Gopal. 1986. “A Perceptual Model of Vowel Recognition Based on the Auditory Representation of American English Vowels.” *Journal of the Acoustical Society of America* 79.4: 1086–100. <https://doi.org/10.1121/1.393381>.
- Thomas, Erik R., and Tyler Kendall. 2007. “NORM: The Vowel Normalization and Plotting Suite.” <http://lingtools.uoregon.edu/norm/>.
- Trager, George L. 1930. “The Pronunciation of ‘Short *a*’ in American Standard English.” *American Speech* 5.5: 396–400. <https://doi.org/10.2307/452819>.
- Van der Harst, Sander, Hans Van de Velde, and Roeland Van Hout. 2014. “Variation in Standard Dutch Vowels: The Impact of Formant Measurement Methods on Identifying the Speaker’s Regional Origin.” *Language Variation and Change* 26.2: 247–72. <https://doi.org/10.1017/S0954394514000040>.
- Wagner, Suzanne Evans, Alexander Mason, Monica Nesbitt, Erin Pevan, and Matt Savage. 2016. “Reversal and Re-organization of the Northern Cities Shift in Michigan.” *University of Pennsylvania Working Papers in Linguistics* 22.2: 171–9. <http://repository.upenn.edu/pwpl/vol22/iss2/19>.
- Wassink, Alicia Beckford. 2006. “A Geometric Representation of Spectral and Temporal Vowel Features: Quantification of Vowel Overlap in Three Linguistic Varieties.” *Journal of the Acoustical Society of America* 119.4: 2334–50. <https://doi.org/10.1121/1.2168414>.
- Wong, Amy Wing-mei. 2012. “The Lowering of Raised-THOUGHT and the Low-Back

Distinction in New York City: Evidence from Chinese Americans.” *University of Pennsylvania Working Papers in Linguistics* 18.2: 157–66.

<http://repository.upenn.edu/pwpl/vol18/iss2/18>.