UNIVERSITY OF BIRMINGHAM

University of Birmingham
Research at Birmingham

Bring on the crowd! Using online audio crowdsourcing for large-scale New England dialectology and acoustic sociophonetics

Chaeyoon, Kim; Reddy, Sravana; Stanford, James; Wyschogrod, Ezra; Grieve, Jack

Document Version
Peer reviewed version

[Link to publication on Research at Birmingham portal](#)

Bring on the crowd! Using online audio crowdsourcing for large-scale New England dialectology and acoustic sociophonetics

Chaeyoon Kim[*], Sravana Reddy[*], James Stanford[*], Ezra Wyschogrod[**], Jack Grieve[***]

Dartmouth College[*], Georgetown University[**], University of Birmingham[***]

Abstract: This study aims to (i) identify current geographic patterns of major New England dialect features relative to prior work in previous generations and (ii) test the effectiveness of large-scale audio crowdsourcing by using Amazon Mechanical Turk to collect sociophonetic data. Using recordings from 626 speakers across the region, this is the largest acoustic sociophonetic study ever conducted in New England. Although face-to-face fieldwork will always be crucial in sociolinguistics and dialectology, we believe that online audio-recorded crowdsourcing can play a valuable complementary role, offering an efficient way to greatly increase the scope of sample sizes and sociolinguistic knowledge.

Keywords: dialectology, dialect geography, sociophonetics, New England, online crowdsourcing, vowels

Corresponding Author:
James N. Stanford
6220 Reed Hall
Dartmouth College
Hanover, NH 03755
(603)646-0099
Email: James.N.Stanford@Dartmouth.edu

Bring on the crowd! Using online audio crowdsourcing for large-scale New England dialectology and acoustic sociophonetics

Abstract: This study aims to (i) identify current geographic patterns of major New England dialect features relative to prior work in previous generations and (ii) test the effectiveness of large-scale audio crowdsourcing by using Amazon Mechanical Turk to collect sociophonetic data. Using recordings from 626 speakers across the region, this is the largest acoustic sociophonetic study ever conducted in New England. Although face-to-face fieldwork will always be crucial in sociolinguistics and dialectology, we believe that online audio-recorded crowdsourcing can play a valuable complementary role, offering an efficient way to greatly increase the scope of sample sizes and sociolinguistic knowledge.

1. Introduction

New England is a fascinating location for online dialect geography since it has some of the smallest and most fine-grained dialect sub-regions in North America. The early English colonies left a long-term sociolinguistic footprint on this region, namely, the Founder Effect, whereby social and linguistic patterns of the earliest founding settlements persist for many generations (Wolfram & Schilling 2016: 29, 97-110; Mufwene 1996, 1999; Zelinksy 1992). Many New England fieldwork studies have described dialect features and geographic patterns in this region, starting with the 1930s fieldwork of the *Linguistic Atlas of New England* (LANE 1939-1943; Kurath 1939), and continuing in Laferriere (1979), Carver (1987), Nagy (2001), Boberg (2001), Thomas (2001), Labov, Ash & Boberg (2006), Nagy & Roberts (2004), Roberts (2006, 2007, 2016), Dinkin (2009), Villard (2009), Nagy & Irwin (2010), Johnson (2010), Wood (2011), Ravindranath (2011), Stanford, Leddy-Cecere & Baclawski (2012), Chartier (2013), Stanford, Severance & Baclawski (2014), Johnson & Durian (2017), and others. In the present study, an online dialect geography project, we aim to identify the current status of New England dialect distinctions and to test online crowdsourcing as a way to collect audio recordings from large numbers of speakers.

Dialectology proceeds according to the technology of each time period. The 1930s fieldwork of the LANE project included painstaking manual transcriptions of over 400 New England informants, as well as low-fidelity audio voice samples of over 290 informants using early audio equipment mounted in a truck (see Purnell 2012).[1] The 1960s fieldworkers of the *Dictionary of American Regional English* (DARE) conducted lexical surveys with 215 New Englanders, and tape-recorded 164 New Englanders with voice samples of varying lengths (Cassidy 1985:lxxxvi-cli). Moving to the modern era, *The Atlas of North American English* (ANAE, Labov et al. 2006) collected audio recordings from 762 participants across North America (Figure 1). The ANAE used the technology of its time: long-distance personal telephone calls and then manual vowel formant analyses of 432 North American participants, including 23 New Englanders.[2]

---

[1] Some of these 1930s-era LANE informants' voice samples are very brief; other samples are 20 minutes or longer in length. See Johnson & Durian (2017).

[2] We find 23 New Englanders in the ANAE/Telsur acoustic analysis spreadsheet. The ANAE maps show that an additional 30+ New Englanders were analyzed as part of the auditory/perceptual coding of mergers, etc.

But many new technological tools have emerged over the past decade since the ANAE. In the present study, we take advantage of the technology of our era in an effort to produce the largest-ever acoustic sociophonetic study of New England. We used large-scale online crowdsourcing to collect and analyze 626 New England audio recordings (Figure 2), along with 535 self-reporting questionnaires. Specifically, we used the crowdsourcing site Amazon Mechanical Turk to elicit audio recordings from individual speakers in a way that requires far less time and labor than traditional face-to-face interviews or telephone calls. In addition, we processed the data using computational semi-automated alignment and formant extraction, another method that was not available in the prior eras. Personal contact between researchers and participants, such as face-to-face fieldwork or personal telephone calls, will always be a crucial part of sociolinguistics and dialectology. But we believe that online audio crowdsourcing and other recent technological advances can greatly expand the scope of research and serve as a valuable complementary source of sociolinguistic knowledge.
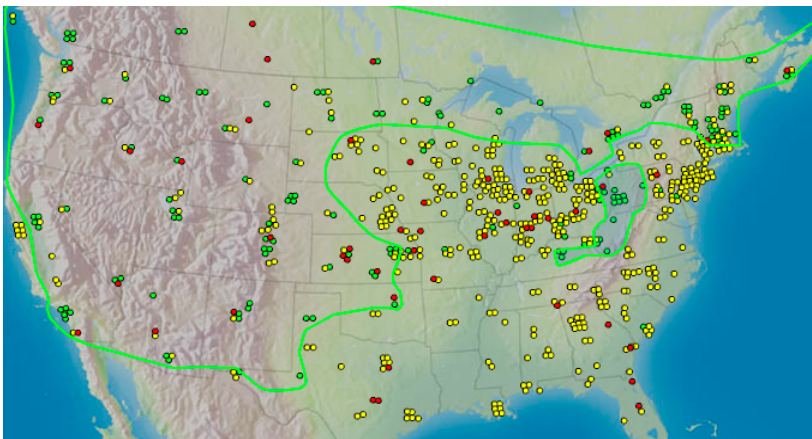
Figure 1. ANAE map of North America based on 762 telephone interviews, collected over approximately nine years from 1991-2000 (Labov et al. 2006: 21). Green lines indicate regions with the low-back merger, discussed in section 5 below. [We will ask for publisher permission for all reprinted maps in this paper.]
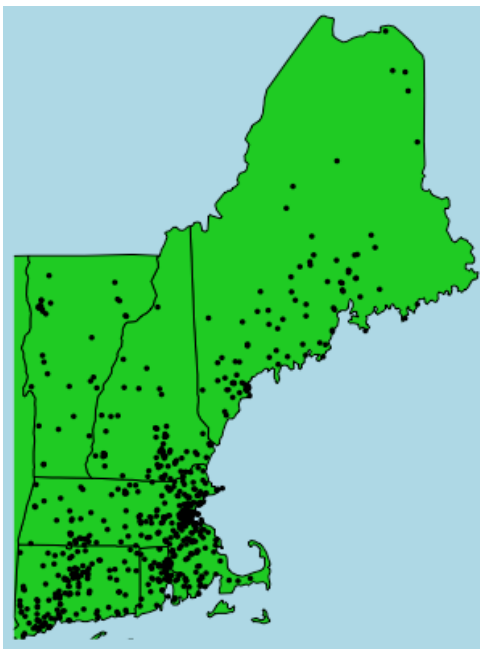
Figure 2. Present study: New England map of the 626 speakers in our Amazon Mechanical Turk audio recording project, collected online over four months from September-December 2016, mapped according to speakers' childhood homes.

Our Mechanical Turk data points provide a reasonably good representation of New England and its population areas. Comparing Figure 2 with the US Census map in Figure 3, we observe that the distributions of our Mechanical Turk data points generally reflect the census population densities across New England. Naturally, there are fewer Mechanical Turk respondents from small rural towns than from the highly populated urban areas (eastern Massachusetts, southeastern New Hampshire, Rhode Island, New Haven and Hartford, Connecticut, etc.). Nonetheless, in our Mechanical Turk study, many rural regions are reasonably well represented for their sizes.
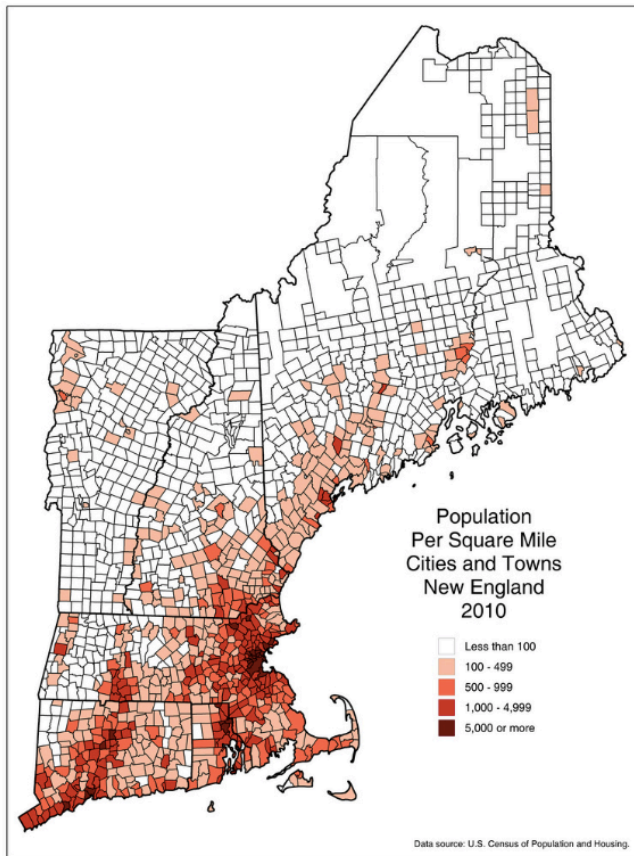
Figure 3. New England by population density, based on US Census 2010 data.
(https://c1.staticflickr.com/9/8328/8103469903_2cd83fda61_b.jpg)

2. Prior work

A detailed analysis of prior work on New England would go beyond the scope of the present paper,[3] but we briefly review some of the major variables being considered in the present study. Figures 4-5 show the traditional New England dialect regions, originally delineated by LANE (Kurath 1939, 1949), and

---

[3] A full literature review or technical discussion of each of these New England variables would greatly expand the paper beyond its focus (the Mechanical Turk project), so we hope that interested readers will consult the prior literature cited in this section for more information about particular variables.

4

then largely confirmed with Carver's (1987) lexical analysis of DARE. We note the general agreement of an east-west contrast running along the Green Mountains of Vermont and southward into western Massachusetts. Kurath defined eastern New England in terms of an isogloss bundle that includes r-lessness, "broad-a" BATH vowels,[4] fronted START/PALM vowels, and other features. Also notable in Figure 4 is Carver's contrast between Rhode Island and Massachusetts.

a. From *LANE* (Kurath 1949, fig. 3; Kurath and McDavid 1961, map 2)
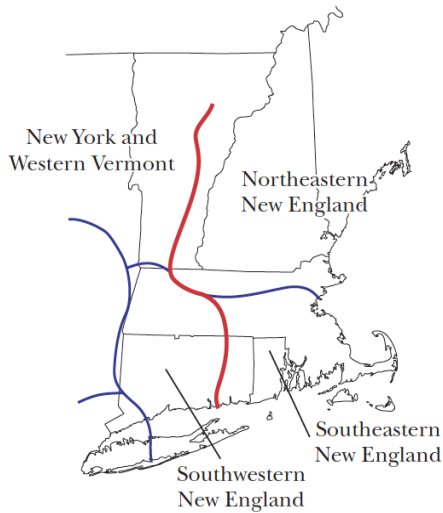
b. From *DARE* (Carver 1987, 31)
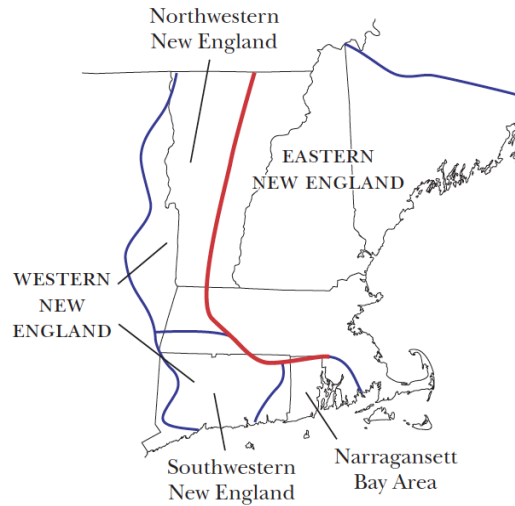
Figure 4. Major New England dialect boundaries in LANE (left) and DARE (right). Maps adapted by Charles Carson and reprinted from Stanford et al. (2012).

---

[4] For convenience in comparing across different diverse dialects of English, we use Wells' (1982) lexical sets in this paper. For some vowels, we find it necessary to use additional subsets not found in Wells, such as SHOUT versus LOUD as subsets of Wells' larger MOUTH set, in order to indicate voicing contrasts and other relevant contrasts in this region.
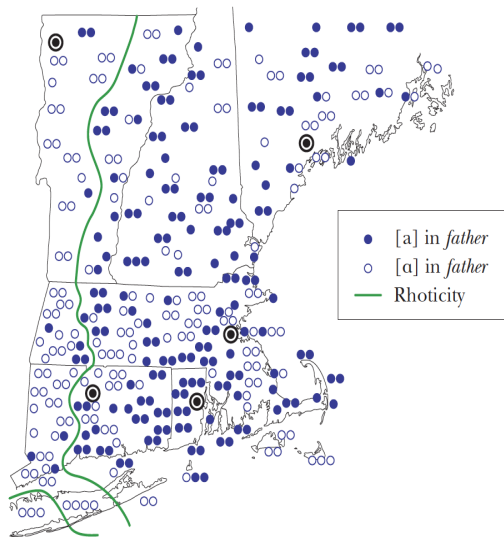
Figure 5. Rhoticity and PALM-fronting (*father*) in LANE. Map adapted by Charles Carson, based on Kurath and McDavid 1961 map 32, reprinted from Stanford et al. (2012).

The ANAE (Labov et al. 2006) also reports an east-west contrast such that Eastern New England (ENE) is characterized by r-lessness, PALM-fronting, and other features. The ANAE (p. 232) finds nasal split short-a (BAT versus BAN) across most of New England, and a sharp geographic contrast in the low-back merger (p. 229): Rhode Island, Connecticut, and western Massachusetts are unmerged, while the rest of New England is merged. Johnson (2010) provides a thorough and fine-grained analysis of this contrast along the Rhode Island/Massachusetts border.

Other related work includes Dinkin's (2005) study of the MARY/MARRY/MERRY distinction in LANE data. Nagy (2001) uses a questionnaire to examines contrasts in PALM/LOT and MARY/MARRY/MERRY between Massachusetts and different parts of New Hampshire, while Roberts (2006, 2007, 2016) explores t-glottalization and /au/ and /ai/ diphthongs in Vermont. Stanford et al. (2012) and Stanford et al. (2014) find that many traditional Eastern New England features are receding among younger speakers in northern New England. Nagy & Irwin (2010) examine rhoticity in Boston and southeastern New Hampshire. Recent fieldwork in Boston neighborhoods (Sipple et al. 2015; Browne & Stanford 2017) finds that speakers raised in traditionally working-class neighborhoods (South Boston, Dorchester, Quincy, etc.) are maintaining many traditional ENE features, while speakers in Greater Boston suburban middle- and upper-class neighborhoods show a rapid decline of ENE features in apparent time. A study of African American communities in traditional Boston neighborhoods (Dorchester, Hyde Park), found speakers to be conservative in some features such as the MARY/MARRY/MERRY distinction but less likely to participate in START-fronting or PALM-fronting (Browne & Stanford 2017). Moreover, the African American speakers in that study tended to have distinct LOT/THOUGHT, which differs from the LOT/THOUGHT pattern for eastern Massachusetts discussed above.

In view of the variables in the literature reviewed here as well as other New England analyses, the present study examines (non)rhoticity, fronted START, fronted PALM, LOT/THOUGHT, MARY/MARRY/MERRY, raised PRICE, raised SHOUT, and nasal split short-a.

6

3. Methods

Amazon Mechanical Turk (https://www.mturk.com) is an online crowdsourcing marketplace where individuals and businesses can upload simple tasks for workers to complete, and the workers receive a small monetary compensation in return, typically ranging from a few cents to a few dollars. For this study, we sent out two waves of surveys; the first was a self-reported questionnaire on lexical and phonological variables by users in the US Northeast area, and the second was an audio recording project focused exclusively on the six New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont). Other researchers have used online methods for sociolinguistics, but typically in questionnaire formats such as Vaux & Golder's Harvard US Dialect Survey (2003), and Vaughn & Kendall's (2016) Mechanical Turk perceptual study of (ING). Wood, Horn, Zanuttini & Lindemann (2015) show how questionnaires in Mechanical Turk can be used to explore syntactic microvariation.[5] We believe that the present study is the first time that Mechanical Turk has been used to elicit large-scale *audio* recordings that can be analyzed in terms of sociophonetics and dialectology. Moreover, prior online work has tended to collect data across entire nations like the US, whereas our audio study focuses on fine-grained regional detail in New England, emphasizing local dialect features in the field materials. Other online audio work includes Bowern's (2010) web-based collection of word-list recordings across the U.S. (which we credit as generating the idea for our study), and two studies using phone apps as dialectology tools for British English (Britain, Leemann, & Kolly 2016) and Swiss German (Leemann, Kolly, Purves, Britain & Glaser 2016).

Overview of the two Mechanical Turk projects

As noted, this study involves two different Mechanical Turk (MT) projects: (1) a self-reporting questionnaire project and (2) an audio recording project where speakers recorded themselves reading. Due to space considerations, we only analyze phonological features in this paper, not lexical items. But we note that the questionnaire produced useful results demonstrating certain regional lexical differences as well, such as the regional usage of the word *jimmies* to represent ice cream sprinkles, *packie* or *package store* for liquor stores, and *wicked pissah* to mean 'very awesome.' The self-reporting questionnaire consisted of questions about 35 regional lexical items and 7 phonological features: rhoticity, intrusive-r, PALM/ LOT/THOUGHT, MARY/MARRY/MERRY, NORTH/FORCE, and "broad-a" BATH (see Appendix B). For the self-reported questionnaire, we collected 535 Mechanical Turk responses in less than a month. We paid 40 cents per questionnaire, which seemed to be a typical compensation for a questionnaire of such length. While the audio project focused on the six New England states to maximize the level of New England detail in the results, the self-reporting questionnaire included the New England states plus four nearby states (New York, New Jersey, Pennsylvania, and Maryland). The nearby non-New England states were included in the self-reporting questionnaire simply because we wanted to test whether some lexical items posited to be distinctive to New England actually extended farther into nearby states.

In the audio project, we collected over 800 audio-recorded responses in four months. We required all speakers to be adults who were raised in New England: "To participate in this survey, you must be over 18 years old and have spent most of your childhood in one of these states: Vermont, New Hampshire, Maine, Connecticut, Rhode Island, or Massachusetts" (see Appendix A). In this way, we only accepted data from adult speakers who reported that they were raised in New England. We recognize that this form of qualifying participants relied on the participants' subjective interpretations of the words 'most' and 'childhood'. Future studies could attempt a more precise formulation.

---

[5] See also Burnett (2012) for the role of microvariation in semantics.

The speakers were asked to read 12 sentences twice and upload the recordings to our site. Each speaker was given $2-$4 dollars for their efforts.[6] The sentences were designed to be easily readable and to include 97 targeted vowel tokens in stressed positions. Each sentence was read twice, so we aimed to collect 194 of the targeted tokens per speaker. A handful of speakers failed to read the instructions properly and only read some of the sentences once. In most cases, we were able to contact speakers to ask them to re-record their sentences if they omitted sentences or uploaded a blank recording or failed to read the sentences twice, etc. Mechanical Turk allows the task manager to withhold payment until a user has properly completed the task.

For the audio project, the specific sentences and the targeted word classes are listed below:
1) I hope that Mary bought coffee and pizza at the food shop.
2) Sue rode a tan horse to the farm. The horse likes to kick my foot.
3) This old bus can easily carry the beanbags and the laundry bin.
4) I guess that Sherry didn't bother to start my car or lock my bike.
5) My father sometimes hides his boots by the road in the park.
6) Larry said "I do!" and then he took the candy heart from my palm. I will marry him.
7) Steve tried to shout calmly, "Hey! I thought that Mary paid for the boarding passes!"
8) Pat laughed and laughed at the merry sound of the shouting. The hoarse voice really sounded like Todd.
9) Joe tossed five books into Mary's room – one at a time.
10) In this hot, sunny weather, I could fall down at the drop of a feather.
11) My job has taught me to be calm, say "thank you," shake hands, and talk quietly.
12) I doubt that my father actually bought a very nice card.

Targeted word classes in the audio project:
- START: *start, farm, car, park, heart, card* – 12 tokens (6 tokens x 2 times each)
- PALM: *palm, father (x2), calm, calmly* – 10 tokens
- LOT: *shop, bother, lock, Todd, drop, job* – 12 tokens
- THOUGHT: *thought, bought (x2), coffee, laundry, tossed, fall, taught, talk* – 18 tokens
- ASH – 26 tokens:
  o ASH followed by nasal: *tan, candy, hands, thank* – 8 tokens
  o ASH not followed by nasal: *laughed (x2), passes, Pat, actually, beanbags* – 12 tokens
  o BATH: *laughed (x2), passes* – 6 tokens
- NORTH: *horse (x2)* – 4 tokens
- FORCE: *hoarse, boarding* – 4 tokens
- MARY: *Mary, Mary's* – 4 tokens
- MERRY: *merry, Sherry, very* – 6 tokens
- MARRY: *marry, Larry, carry* – 6 tokens
- "Canadian Raising"
  o PRICE: *bike, like, likes, nice* – 8 tokens
  o PRIDE: *hides, tried, time, five* – 8 tokens
  o SHOUT: *shout, shouting, doubt* – 6 tokens

---

[6] We began the audio project with a pay rate of $2 per person, but later decided to increase the pay rate to $4 in an effort to slightly accelerate the number of daily responses. In fact, however, the number of daily responses was approximately the same regardless of the change in pay rate.

      o LOUD: *sound, sounded, down* – 6 tokens
- Rhoticity: *horse (x2), farm, bother, start, car, father (x2), park, heart, boarding, hoarse, weather, feather, card* – 30 tokens

We recognize that this reading activity can only provide data from a single, relatively formal speech style, but we believe that the results (section 5) nonetheless provide valuable patterns for analysis in this speech style. Future studies could try free speech activities as well, although this will naturally increase the time needed for processing. In that case, much greater time resources for manual annotation would be required because, like other semi-automated vowel alignment and extraction systems, DARLA requires a human annotated transcript as input[7]. In the present study, we asked the participants to read informally: "Please use your normal everyday voice, not deliberately polished or formal" (see Appendix A).

Collecting demographic information
Figure 6 shows the first page of our survey where we collect demographic information. This figure shows how respondents provide their primary childhood location ("During ages 0-12, what is the name of the city/town you spent the most time in?" Etc.). Subsequent pages ask for the primary location of their teenage years (13-18 years old) and adulthood.

Figure 6. The first page of our Mechanical Turk demographic information section.

---

[7] DARLA also has a completed automated feature where the system uses automatic speech recognition (ASR) to transcribe the words, without any human intervention. But since publicly available ASR systems are not very accurate, this completely automated option is far less accurate. For accurate vowel measurements as needed in this study, we used DARLA's semi-automated functions, not the completely automated functions.

In the data analysis stage, we converted each participant's childhood zip code or town into geographic coordinates (latitude/longitude). To increase the granularity in urban areas, our survey asked participants from larger cities to specify a neighborhood (e.g. Bronx, NY in lieu of "New York, NY" or Brighton, MA instead of "Boston, MA.")

The audio uploading interface
Mechanical Turk is commonly used for human subject research in psychology and other fields, but it is not yet widely used in linguistics. One of the challenges that we faced with Mechanical Turk was to set up an interface where the workers could upload their recordings. Currently, Mechanical Turk does not offer direct ways to upload files, so we created our own form on a Dartmouth website and linked the form to the task on Mechanical Turk.

Microphone quality
In the audio project, we had to depend on the microphones and computers owned by the respondents, which naturally leads to some uncontrolled factors in the quality of the recordings. In general, however, we find that most computers nowadays have fairly consistent microphone qualities. Even poor computer microphones of today have better fidelity and audio range than the telephones of the ANAE era, not to mention the highly outdated recording devices used in LANE and DARE. As discussed below, we took many precautions to ensure that our analysis only included reliable recordings, including manually listening to every sentence from every speaker, removing noises where possible, and completely omitting speakers with poor-quality recordings. We also note that many published studies have compared vowel formants in modern field recordings with legacy recordings (such as LANE and DARE), even though the differences in recording quality across those data sets is even greater than the differences between our respondents' home computer microphones. For example, Thomas' (2010) study of Ohio uses vowel formants extracted from modern recordings and also DARE 1960s recordings. Likewise, Labov, Rosenfeld & Fruehwald (2013) compare vowels across decades of Philadelphia field recordings, a generational project which presumably includes a wide range of different microphones and recording devices over many years.

Manual checking of each recorded sentence
832 speakers uploaded their recordings on our Mechanical Turk site during the four months of the project (September-December 2016). We found it necessary to omit 90 of the speakers due to background noise,[8] sound quality, major reading errors, etc. We carefully listened to each sentence recorded by each speaker. This process of manually checking included quality control for minor reading errors, background noise, and any other problems. Many minor issues could be manually corrected in Praat, such as deleting extraneous noises, laughter, or other voices, as well as repairing some minor reading errors to match our transcriptions. We conducted acoustic analysis on 626 speakers,[9] at which point we had a gender proportion of 55% women (N=342) versus 45% men (N=284). We stopped processing the dataset at this point because we wanted to avoid letting the gender imbalance become greater than 55%

---

[8] We tried noise reduction software in some cases, but we did not find it to be helpful for the purposes of formant tracking. Although such software reduces noise, it often gives the recording a highly "muffled" or otherwise artificial sound, and does not appear to improve formant tracking in a significant way.

[9] One speaker did not give a birth year, so that speaker is omitted in analyses involving birth year, leaving 625 in those cases.

women to 45% men; we therefore left 116 speakers unanalyzed. For whatever reason, it turns out that more women than men chose to participate in our Mechanical Turk audio study.[10]

Alignment, extraction, and normalization

After each recorded sentence of each speaker was manually checked by human analysts, we used DAR-LA (Reddy & Stanford 2015) to align the phonetic segments and extract the vowel formants. DARLA, which stands for "Dartmouth Linguistic Automation," is an alignment and extraction system that uses FAVE-Extract (Rosenfelder et al. 2014), Prosodylab-Aligner (Gorman, Howell & Wagner 2011), and the Vowels R package (Kendall & Thomas 2010). Computational alignment and vowel extraction methods have been implemented and tested in a number of previous studies (Evanini 2009:92; Evanini, Isard & Liberman 2009:3-4; Labov et al. 2013; Stanford et al. 2014; Severance, Evanini & Dinkin 2016; Stanley & Renwick 2016). In addition, we randomly selected 10 aligned TextGrids from the DARLA output and manually checked them in Praat, and found the alignments to be reliable.

We recognize that the scale and level of automation in this project requires a greater tolerance of error than traditional methods that use less automation, but we believe that it is a reasonable tradeoff for the benefits of a large dataset. Naturally, different types of studies have different goals, strengths, and weaknesses. A smaller-scale study of 30-40 speakers would make it possible for researchers to learn more about each speaker's individual opinions and circumstances, and perhaps attain a higher level of accuracy in the vowel analysis. But on the other hand, smaller studies are more limited in their ability to reliably represent a large population as a whole. Moreover, as Fruehwald (2014) points out, automation like FAVE and DARLA allows for greater replicability and data-sharing/direct comparisons between researchers' datasets.

FAVE-Extract extracts F1 and F2 at a single point in each syllable determined by the heuristic established in the ANAE (the specific criteria are explained in Labov et al. 2006:38 and Labov et al. 2013:35-6). At that point in the syllable, FAVE-Extract automatically tests for the optimal number of poles (nFormants) to select the best F1/F2 measurement for each token. The output also provides F1/F2 measurements at 15% increments along the syllable, although the present study just examines basic F1/F2 space as extracted from the single FAVE-Extract point. An analysis of other variables, including vowel duration or incremental formant values across the syllable, would go beyond the present study, but we look forward to future analyses of these additional aspects of the data.

Using our dataset of 626 speakers, we filtered out vowels in unstressed positions, vowels in grammatical function words, and tokens with unreliable high-frequency bandwidths. In addition to this filtering, we also note that FAVE-extract itself omits tokens where formants cannot be measured reliably, further increasing the accuracy of the dataset in representing a speaker's vowel space. In the final filtered dataset, we had 107,0179 stressed vowel tokens available for analysis. We then normalized in R using the Lobanov method from the Vowels package (Kendall & Thomas 2010), and converted to a Hz scale for convenience in discussion. Figure 7 shows plots of the individual vowel tokens by gender. The graphs demonstrate that the vowels of the female and male speakers have been normalized appropriately in terms of vowel space size.

---

[10] We recognize the value that each speaker's voice contributes to the dataset. However, for the present research questions, if we had included the remaining 116 speakers, the gender balance would have increased from 55/45 to 62/38, which we consider an unbalanced and unrepresentative ratio that would make multivariate analyses more difficult to interpret. The remaining 116 speakers can be analyzed in the future for other research questions.
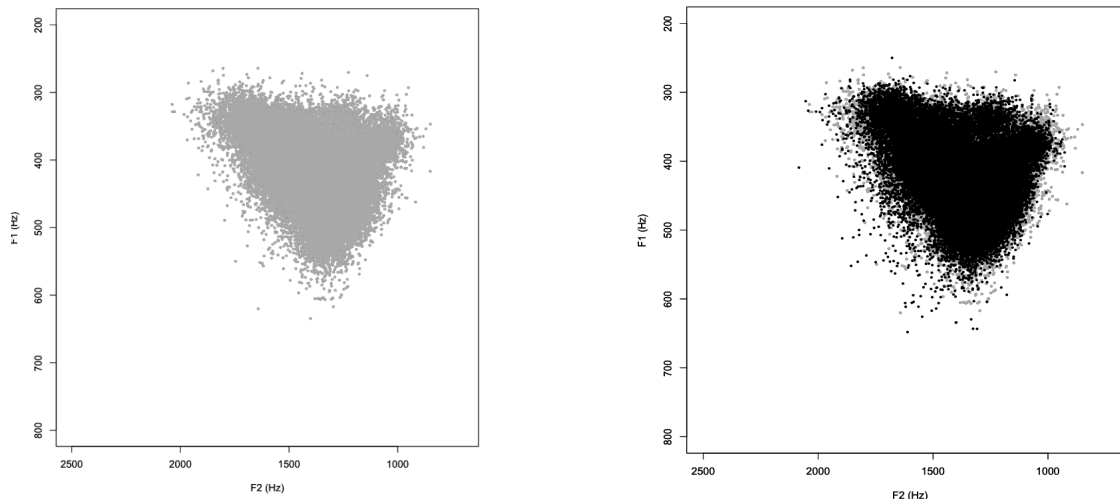
Figure 7. All vowel tokens, organized by gender.
Left: 54,623 tokens from 342 women. *Right:* 52,456 tokens from 284 men.

Reliability of speaker locations

We only focused on childhood locations for the present study. Due to the sheer volume and complexity of this data set, we found it necessary to focus on this aspect of speakers' geographic history. Future work on this dataset can also consider more detailed analyses of mobility by comparing each speaker's childhood locations with teenage and/or adult locations, and so on. Teenage years are an important time in a person's dialect development, and features can change in adulthood as well (e.g., Sankoff 2004, Sankoff & Blondeau 2007, Wagner 2012). Nonetheless, we find that the childhood locations produce effective results for the purposes of the present paper.

Second, as with any online activity, there is always a risk that some people may be misrepresenting themselves. Can we trust data from speakers who have self-identified their geographic locations? After all, we paid these Mechanical Turk respondents, so it is conceivable that some of them are falsifying their information in order to participate. But actually this risk is diminished because Mechanical Turk workers are not even allowed to view our task unless their demographic form shows that they have a New England address. For non-New Englanders to falsify their identities and participate in our task, they would have to (falsely) change their demographic form just for our task. It is plausible that some MT worker made this decision, but we think it would a considerable amount of trouble for just a few dollars, especially since they could earn money more quickly by simply choosing a different task. Moreover, most Mechanical Turk tasks are significantly simpler than our audio recording task (recall that our users had to record themselves 12 times, save each of the 12 recordings separately to a computer, then upload those recordings one at a time). On any given day, Mechanical Turk offers thousands of tasks, many of which are simple surveys and other easy jobs, and so there is little incentive for a person to falsify documentation just to do our (relatively complex) dialect-recording task. Therefore, we expect that the amount of such demographic "noise" is minimal in our data, but we recognize that there is a tradeoff: large-scale data means that there may be more demographic noise.

12

Demographic distribution in the audio project

In terms of demographic distribution, we find that Mechanical Turk makes it possible to reach many pockets of the population that can be difficult to reach in a typical research setting (university campus). In particular, our dataset shows a strong representation of people from lower education and socioeconomic levels. The series of histograms in Figure 8a-e below show the distribution of the workers by education level, occupation level, age, and ethnicity. In an online project like this, we might have assumed that participants would be predominantly college-educated, upwardly mobile professionals. But with Mechanical Turk, it turns out that this is actually not the case at all. Mechanical Turk workers are quite diverse socioeconomically, at least in New England. Our data set includes respondents from educated middle-class backgrounds but also many others with little or no college education and occupations like bus driver, bodyguard, cook, etc., all interested in making a few extra dollars online in their spare time.

The age distribution is not quite as good. As expected for an online study requiring some amount of technical savvy, our age distribution skews toward the younger side (Figure 8c). Even so, middle-aged and older speakers are solidly represented as well. Moreover, the age distribution in Figure 8c is much better than it would be in a traditional campus-based sociolinguistic study limited to 18-22 year-olds.

The weakest aspect of our demographic distribution is the serious lack of ethnic diversity (see Figure 8d), which has been a problem in traditional New England dialectology as well. A recent field-based study helps to improve ethnic representation by means of fieldwork in the African American community around Dorchester and Hyde Park (Browne & Stanford, forthcoming). In Figure 8e we note that our Mechanical Turk ethnic distribution shares some degree of similarity with New England ethnic distributions in the 2010 US Census, but we recognize that non-White speakers are significantly underrepresented in our data.
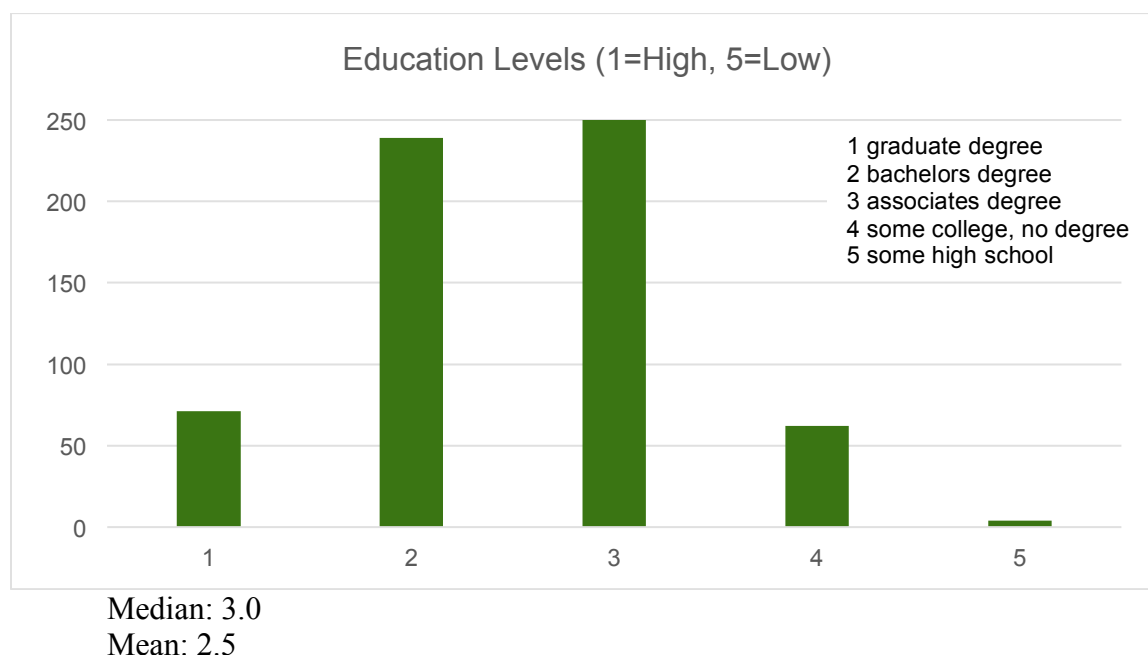

Median: 3.0
Mean: 2.5

Figure 8a. Education levels of the speakers in our Mechanical Turk audio project.

Occupations (1=High, 6=Low)

1 professional, owner-director of large firm
2 white collar – proprietor, manager
3 white collar – merchant, foreman, sales
4 blue collar – skilled
5 blue collar – unskilled
6 unemployed

Mean: 3.0
Median: 3.3
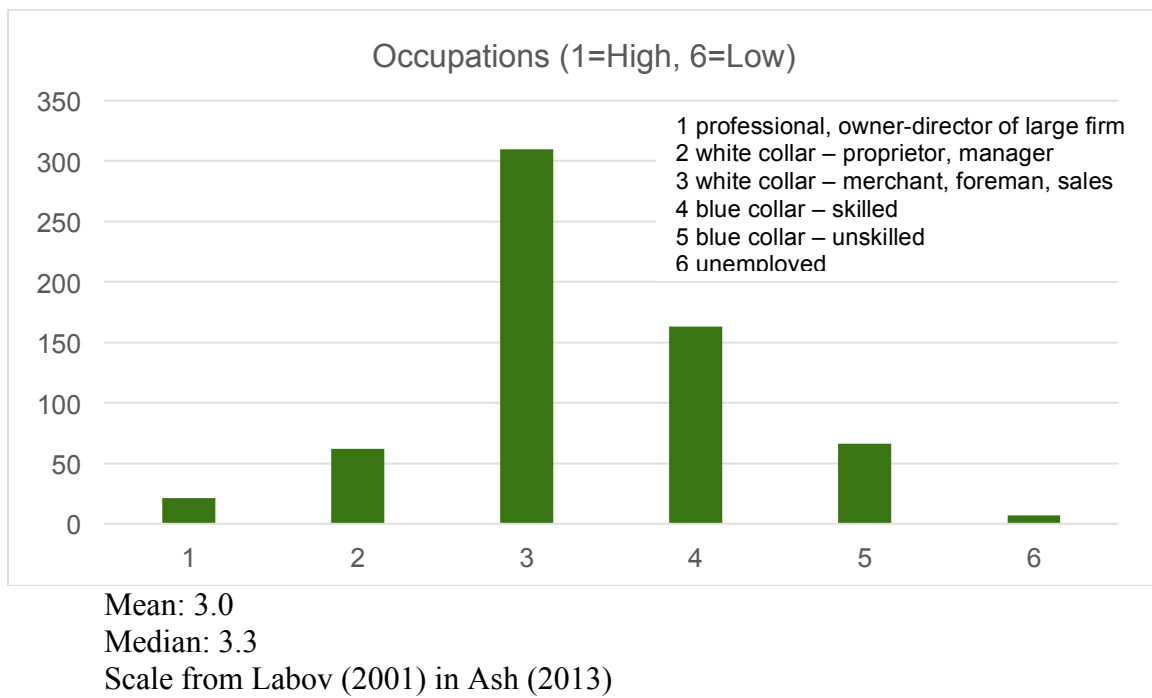Scale from Labov (2001) in Ash (2013)

Figure 8b. Occupation levels of the speakers in our Mechanical Turk audio project.
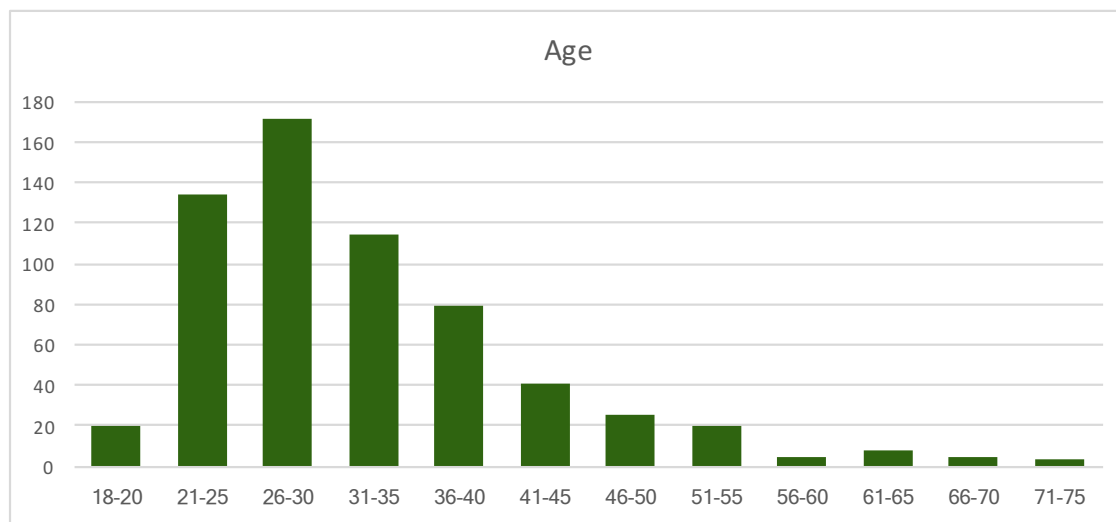

Age

Figure 8c. Age of the speakers, by birth year, in our Mechanical Turk audio project.
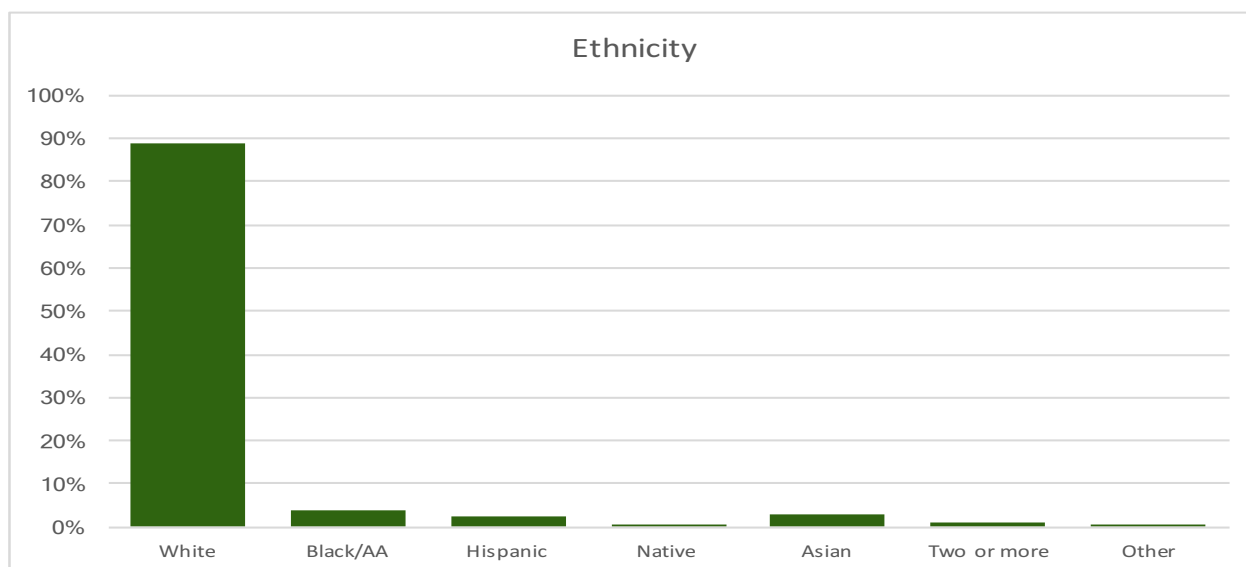
Figure 8d. Self-reported ethnicities of the speakers in our Mechanical Turk audio project.
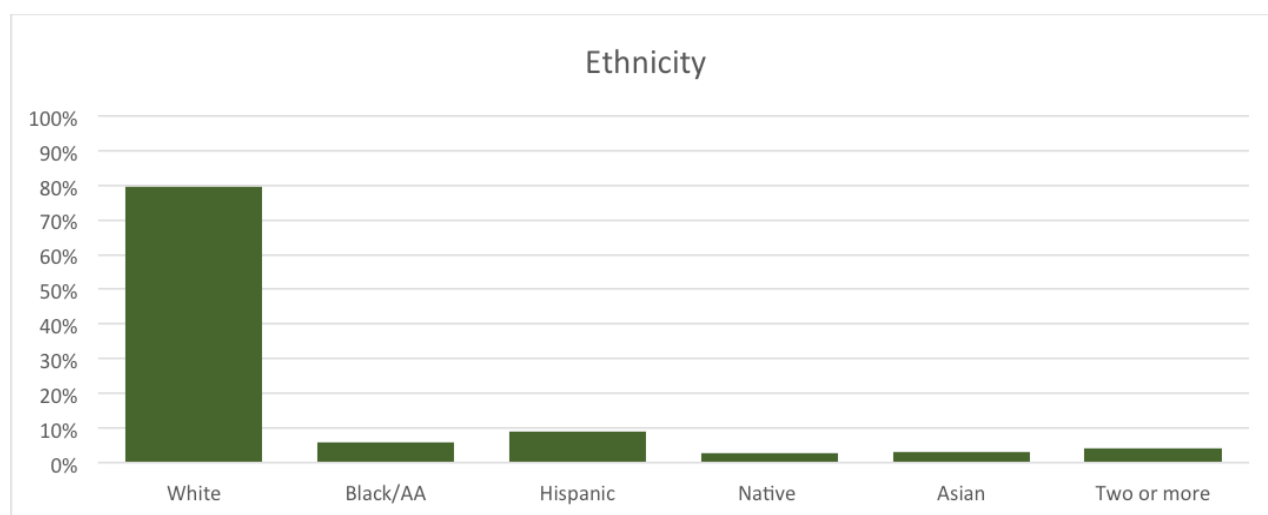

Figure 8e. US Census 2010: New England self-reported ethnicity (%) (based on census counts reported at http://www.nebhe.org/wp-content/uploads/FigDEM061.png)

Archiving and public access
In the consent process, we invited the respondents to give us permission to make their recordings publicly available. The majority of respondents consented for their recordings to be used in this way, and we are currently developing a way to make this dataset publicly accessible. Users will be able to compare New England dialect variables across hundreds of speakers reading the same passages, along with documented demographic and geographic information for each speaker, but no personally identifiable information.

15

5. Results

The results aligned quite well with the regional patterns based on previous studies, suggesting that online crowdsourcing can indeed quickly and accurately differentiate the dialect sub-regions of New England. For each variable, we first briefly show the results of the self-reporting questionnaire and then provide the audio project results (i.e., acoustic analysis) in more detail. We focus on basic dialect geography in the maps, and we also consider other factors in multivariate statistical modeling with Rbrul (Johnson 2009), including age, gender, education, occupation, and regional patterns.

5.1 LOT/THOUGHT

For LOT/THOUGHT, Figures 9-11 provide comparative maps of the ANAE, our self-reporting questionnaire, and our audio project. The isogloss in our MT audio project is remarkably similar to the ANAE: compare Figure 9 with Figure 11. In Figure 11, which is a map of LOT/THOUGHT based on the acoustic analysis of our audio recordings, the shade of the dot represents the Euclidean Distance between the plotted LOT and THOUGHT vowels: darker dots represent smaller Euclidean Distances (LOT and THOUGHT closer together). The mean LOT/THOUGHT distance in Rhode Island, Connecticut, and Western Massachusetts was 75 Hz, whereas the rest of the New England region showed an average of 30 Hz ($p<0.0001$). Considering the consistency between our main isogloss pattern in Figure 11 and the prior work, we believe that these Mechanical Turk-based LOT/THOUGHT results serve as a "proof of concept" for the large-scale online methods and extraction methods in this study. We recognize that Euclidean Distance is a one-dimensional measure that may mask more complex patterns happening in F1/F2 space, as well as differences in duration or voice quality. A full analysis of LOT/THOUGHT as a merger would require perception studies as well.

The Rbrul linear mixed effects analysis (best fit R^2 fixed=0.267, speaker as random effect) modeled the speakers as 0.28 Hz closer for each year younger, 3.2 Hz closer for each step of greater education, 15.9 Hz closer if the childhood location was in the ANAE merged region (cf. Figure 9), and 22.8 Hz closer if followed by a voiced consonant rather than voiceless (see word list in section 3). Occupation and distance from Boston were not significant.
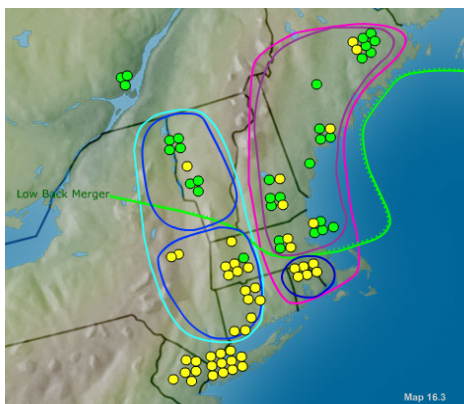

Figure 9. LOT/THOUGHT merger in ANAE (2006). Green=merged. 59 speakers. See also Johnson (2010) for a detailed analysis of the Rhode Island/Massachusetts border.

Do you pronounce the words cot and caught the same or differently?
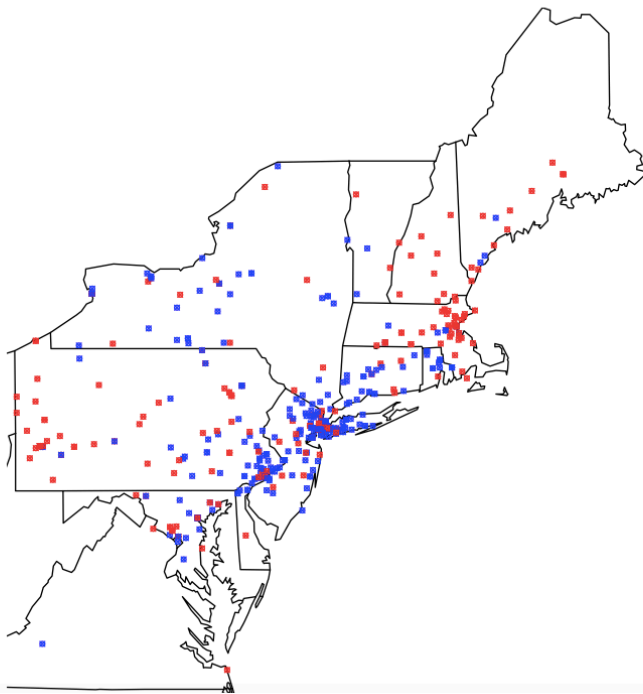**Blue=Same,Black Red=Different**

Figure 10. LOT/THOUGHT map from our self-reporting questionnaire (535 respondents) [wording based on Nagy (2001)]
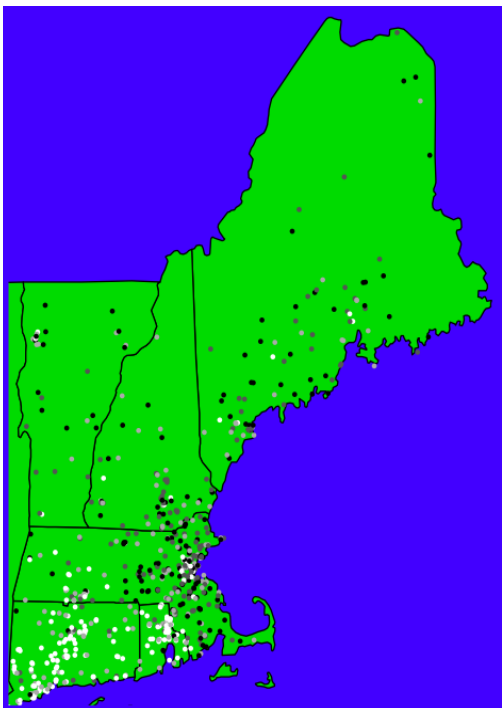
Figure 11. LOT/THOUGHT map based on our 626 speakers in the MT audio recording project. Each dot represents a single speaker. Plotted in quartiles, where darker=closer together (smaller Euclidean Distance between LOT and THOUGHT).

5.2 Rhoticity

Our maps from both the self-reported questionnaire survey (Figure 13) and the audio recording project (Figure 14) align well with maps from ANAE (Figure 12) and other prior work. As Figures 14-15 show, both the self-reporting questionnaire and the audio project showed non-rhotic speech radiating outward from the Massachusetts Bay area. Figure 14 is based on binary auditory judgments[11] of the audio recordings as r-less versus r-ful. For our data points in extreme northern Maine, we note that the proximity to Canada may have a effect as well, so we hope that future fieldwork can provide more insights about this rural far northern Maine region which was not sampled in LANE, DARE, or ANAE.

Modeling with Rbrul logistic regression (Johnson 2009) shows that the log-odds of r-lessness decreases by 0.061 for each later year of birth year (R^2=0.225), and r-lessness decreases by 0.013 for each mile farther from Boston. Log-odds of r-lessness increases 0.406 for each step of higher education, using a 1-5 scale based on Ash (2013:355) following Labov (2001). There was not a significant relationship observed for sex or occupation.
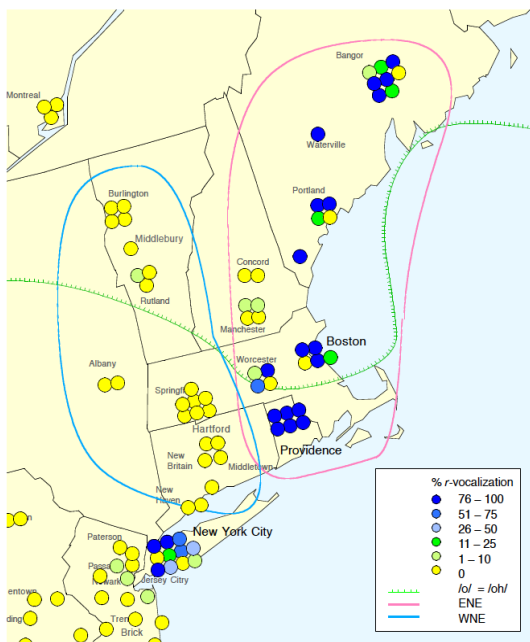
Figure 12. ANAE map of New England rhoticity (Labov et al. 2006)

---

[11] A gradient measure of rhoticity, including a full analysis of phonetic environments, would be more illuminating than these binary counts, of course, but this level of analysis goes beyond the scope and resources of the present study. We hope that such an analysis will be possible in a future study of this large dataset. We are not aware of a reliable fully automated acoustic method that is available for measuring rhoticity yet (Tyler Kendall p.c.; Yaeger-Dror et al. 2009; Heselwood et al. 2008), so we used human auditory judgments for rhoticity here.

Figure 13. Rhoticity map from our self-reporting questionnaire.
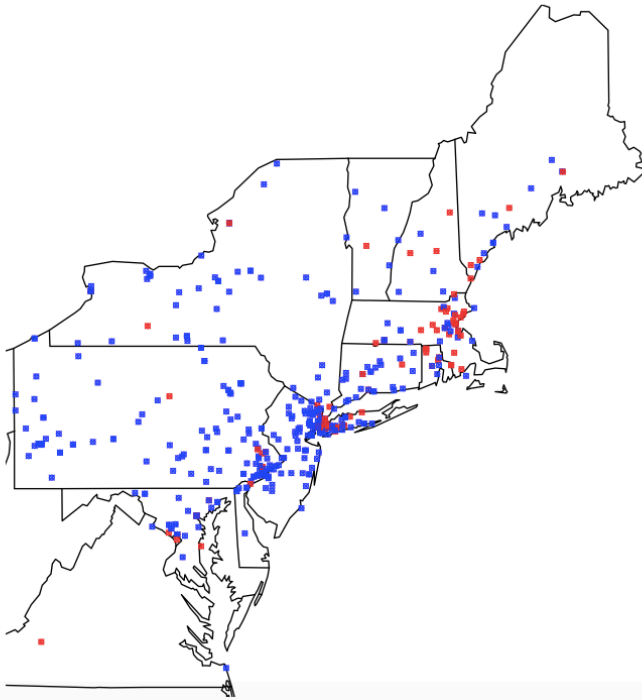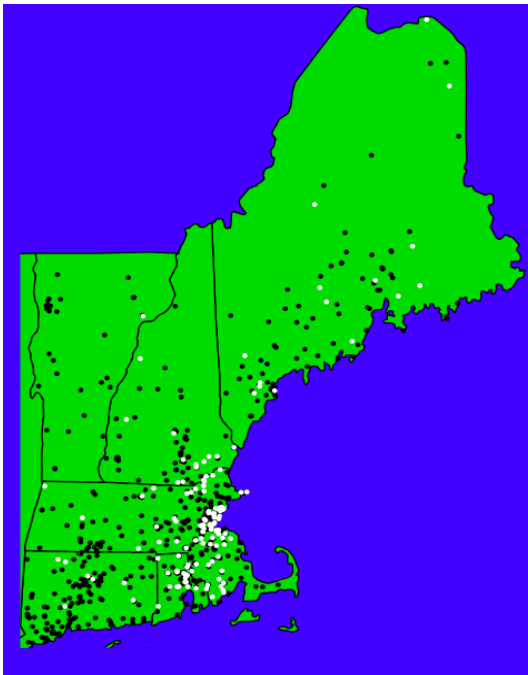

Figure 14. Rhoticity from our MT audio recordings, binary auditory judgments. Light=r-less, dark=r-ful.

5.3 Nasal Split Short-a

As observed in the ANAE's New England analysis (Labov et al. 2006:232), in our MT audio project we find that New England ASH vowels are clearly split according to nasality of the following consonant, with BAN higher than BAT (Figure 15). This effect is stable in apparent time.
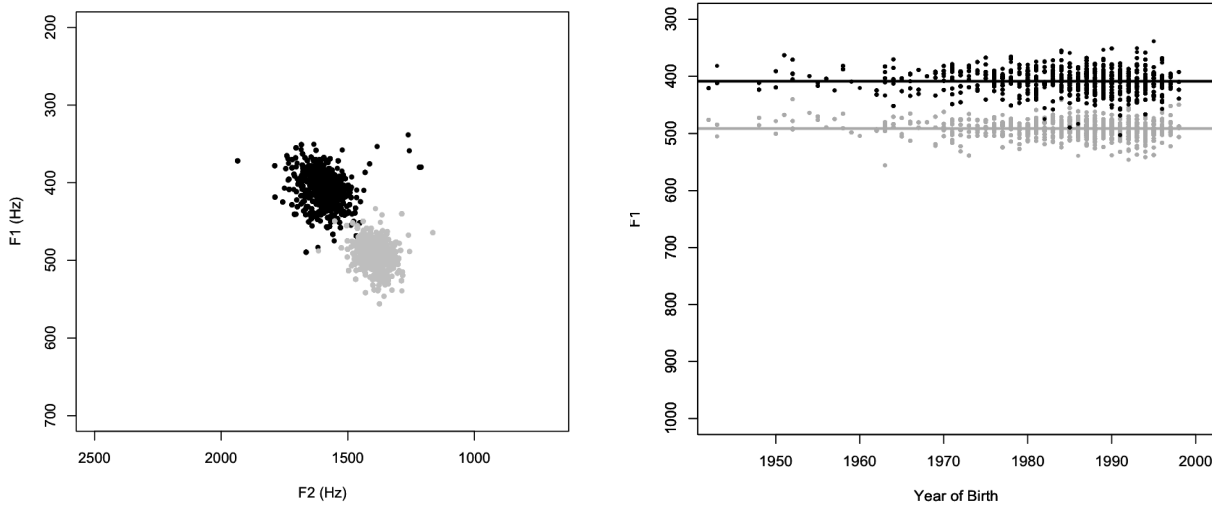
Figure 15. Nasal short-a: Speaker means in our audio data in F1/F2 space (left) and apparent time (right). Dark=followed by nasal (BAN). Gray=non-nasal (BAT).

5.4 "Canadian Raising" in /ai/ and /au/

For "Canadian Raising" in New England (Labov et al. 2006:206; Roberts 2007), our MT audio results show a clear contrast in voicing such that voiceless consonants favor raising of /ai/. This is stable in apparent time. Figure 16 shows the results of our MT audio project for PRICE versus PRIDE.

Figure 16. Speaker means in our audio results for PRICE (black) and PRIDE (gray) in F1/F2 space and apparent time.

However, as expected from the ANAE, such a raising pattern is not observed with the /au/ vowel in SHOUT and LOUD words (Figure 17). Again, this pattern is stable in apparent time.
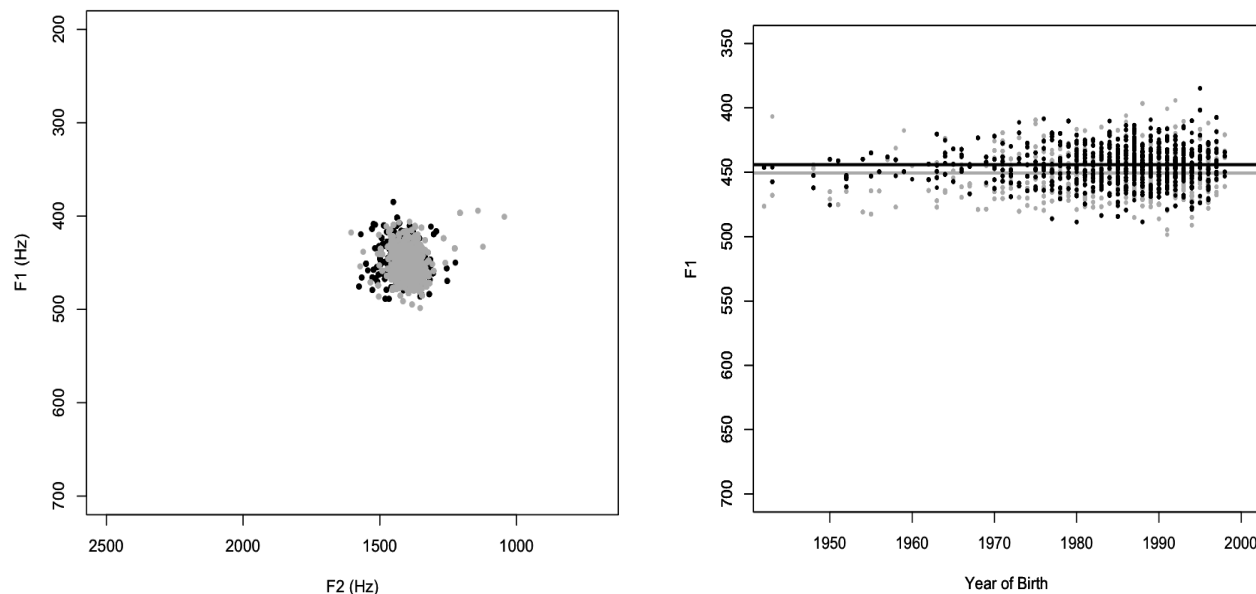
Figure 17. Speaker means in our MT audio results for LOUD (black) and SHOUT (gray) in F1/F2 space and apparent time.

5.5 START and PALM fronting

Following the ANAE approach (p. 231) as reprinted in Figure 18 below, we analyze both START and PALM for possible geographic patterns of fronting 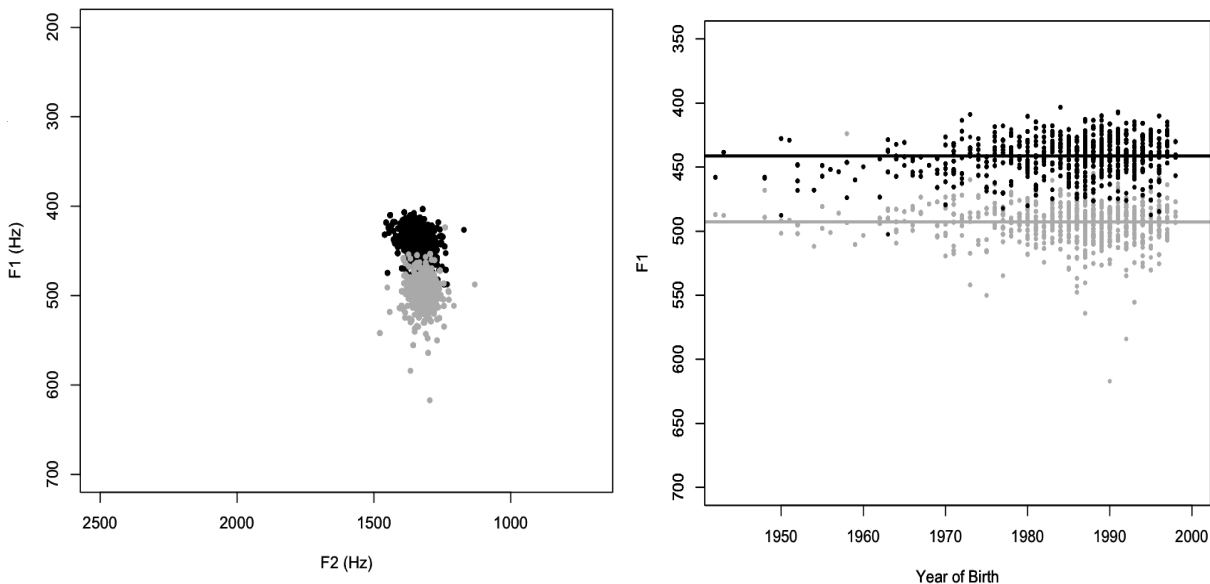in New England. We treat these two vowel classes separately to see how our results compare to the ANAE for these vowel classes. This phonetic analysis is not meant to imply any particular phonemic subcategories of /a/, as this is not a phonology study.

For the MT self-reporting questionnaire, we elicited data on PALM (*When you say father and bother, do they rhyme (like feather and weather)?* [wording based on Nagy 2001]). We did not elicit this for START because we do not have a reliable way to ask for this contrast in START words in a questionnaire with linguistically untrained respondents.

For the audio recording project, the START results are shown in Figure 19. The ANAE (p. 231) analyzes START-fronting in terms of mean START F2 minus mean LOT F2. We take the same approach here so that the results can be easily compared. Our results for START are quite consistent with the ANAE (cf. Figure 18).

Figure 18. ANAE results (reprinted from p.231). Brown dots indicate both PALM-fronting and START-fronting. Red dots indicate START-fronting but not PALM-fronting.
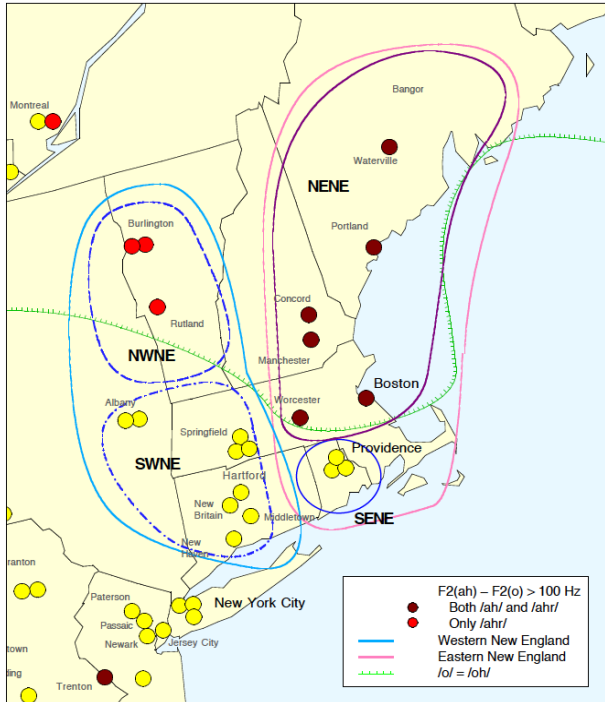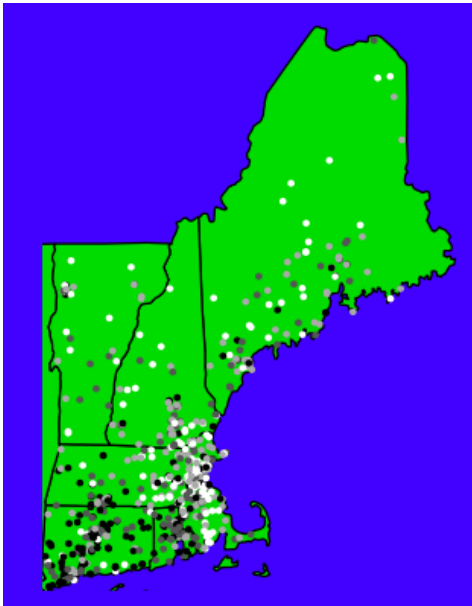

Figure 19. Our MT audio recording results for START-fronting. Plotted in quartiles, where lighter= START vowel is farther forward (higher F2).

As for PALM, in our self-reporting questionnaire we generally see the expected pattern, where non-rhyming *father* and *bother* are generally limited to eastern/northeastern New England and not Rhode Island (compare Fig. 20 and Fig. 18). In addition, note that some respondents from New York City and other non-New England areas also reported that these two words do not rhyme (Fig. 20). In those areas, the reported non-rhyming *father/bother* may be due to backing or raising, rather than fronting. Labov et al. (2006:117) reports PALM-fronting in Eastern New England but backing or raising in New York City. That is, the vowel is perceived as distinct from LOT in both locations, but in different ways. We therefore turn to the acoustic analysis of the audio recordings for greater clarity.
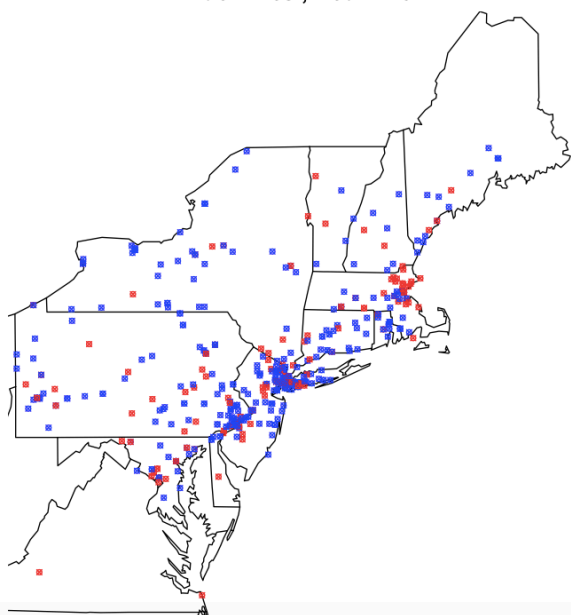
Figure 20. Results of the MT self-reporting questionnaire for PALM: *When you say father and bother, do they rhyme (like feather and weather)?* [wording based on Nagy 2001]). Red = "No", Blue = "Yes".

Figure 21 shows our audio recording results for PALM. Looking at northern New England, we find that PALM-fronting is somewhat less common in Vermont than START-fronting is, which corresponds to the ANAE's observations about PALM versus START in northern New England (cf. Figure 18 above). We do, however, see some sporadic PALM-fronting in parts of Connecticut, which differs from the ANAE results in that area.
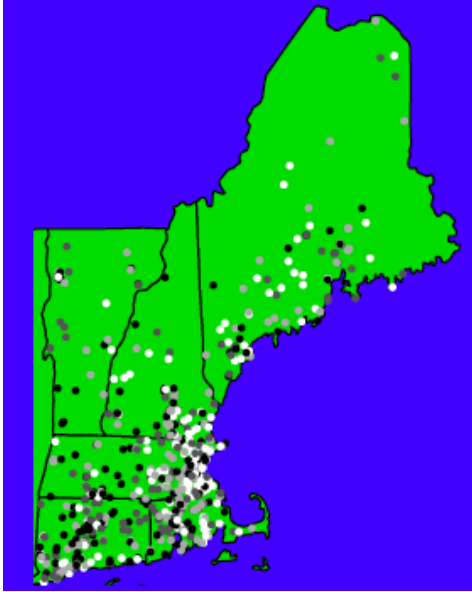
Figure 21. PALM results from our audio recordings, in quartiles, lighter=farther forward (higher F2).

5.6 MARY/MARRY/MERRY

The geographic patterns of MARY/MARRY/MERRY distinctions closely matched prior work. Shown below are the maps from the ANAE (Figure 22), our self-reporting questionnaire (Figure 23), and our audio recording project (Figure 24). For the audio recordings, we calculated the Euclidean "tridirectional" distance between the means of the three vowels for each speaker.
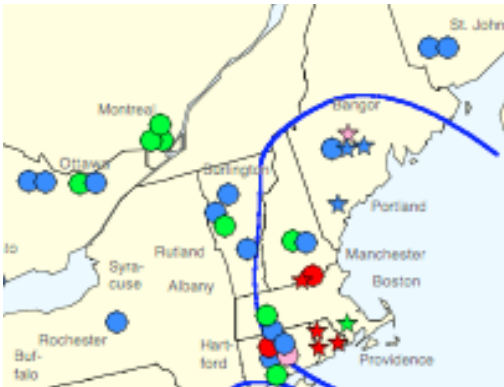
Figure 22. ANAE results for MARY/MARRY/MERRY. Red: all different. Green: two different. Blue: all are the same.
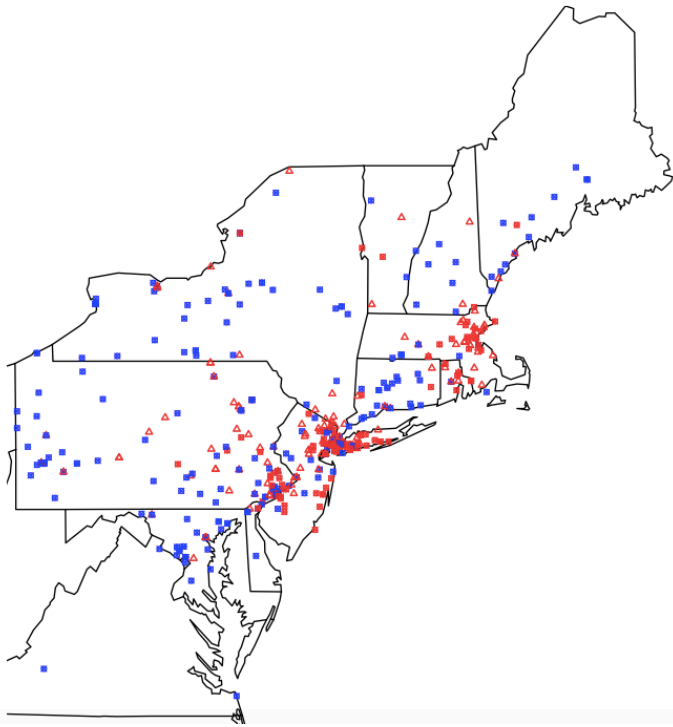
Figure 23. Our MT self-reporting questionnaire results for MARY/MARRY/MERRY. Blue square: all 3 the same, red square: all 3 different, red triangle: 2 of 3 the same.
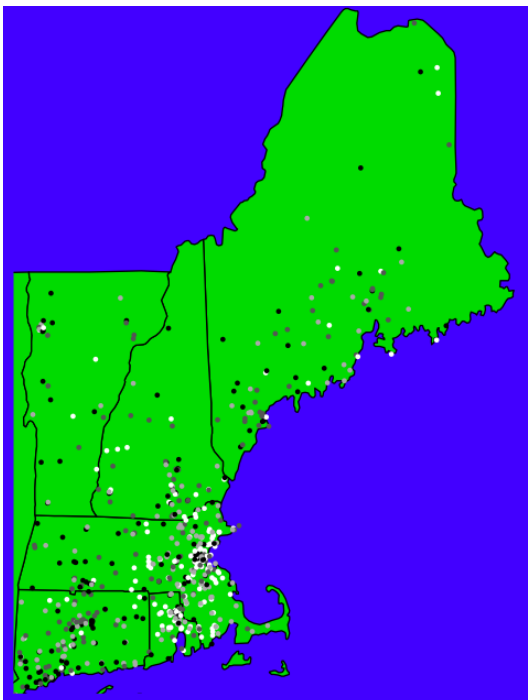
Figure 24. Our MT audio project results for MARY/MARRY/MERRY. Total tridirectional Euclidean distance between the three vowels for each speaker, in quartiles. Lighter = greater distance.

In our Rbrul modeling of the tridirectional distance between MARY, MARRY, and MERRY, the best-fit model has all three vowels of MARY/MARRY/MERRY 4.8 Hz closer together for each year younger, 82.9 Hz farther apart for men, and 15 Hz closer together for every 10 miles farther from Boston (R^2=0.154).

6. Inferring New England Dialect Regions

We conclude this article with an exploration of our Mechanical Turk dataset using ordinary kriging, a common geostatistical technique for interpolation, which has previously been used in dialectology to facilitate the comparison of maps from different dialect studies (Grieve 2015). We conducted this analysis both as a way to visualize the underlying regional patterns exhibited by our five main acoustic variables, and as a precursor to mapping modern New England dialect regions through an aggregated analysis.

Ordinary kriging (Krige 1951; Bivand et al. 2008, Grieve 2017) estimates the values of a variable at unknown locations based on its values at known locations, taking into consideration the associated *variogram* (Atkinson & Lloyd 2009, Grieve 2017), which provides a model of how these values change across space. In this way, ordinary kriging can be used to generate a dialect map that represents the predicted values of the variable across the entire region based on observed values at limited number of locations. In addition, ordinary kriging effectively smooths the map, removing local variability in these observed values so that underlying regional patterns can be visualized. In this context, ordinary kriging is similar to a traditional isogloss analysis. Starting with a dialect map—which is generally characterized by relatively sparse, uneven, and noisy data—ordinary kriging allows for broad regional patterns to be identified and mapped. The advantage of this technique, compared to drawing an isogloss by hand, is that interpolation methods are far more replicable, rigorous, and efficient.

Despite these advantages, we recognize that some analysts prefer to only map raw data points. Our view, however, is that although it is always essential to provide raw maps so that readers can see all the data for themselves, it is often helpful to also provide a smoothed version of these maps, where the underlying patterns can be fully appreciated. This is useful when comparing numerous features, when the geographic patterns are complex or noisy, and when there is an uneven geographic distribution of data points. Interpolating over a consistent set of locations is also especially important when comparing dialect maps or when producing a composite geographic summary of several dialect features that each have their own regional patterns, for example, to identify general areas of transition between dialect regions. This is similar to the way that bundles of isoglosses are used to identify common patterns of regional variation in traditional dialect studies. In this section, we therefore create smoothed maps for each of our five ENE variables and then use these maps to create a single aggregated map, which allows us to plot a border between the Eastern and Western New England dialect regions.

An example from LOT/THOUGHT

To illustrate our application of ordinary kriging, we describe our interpolation of the LOT/THOUGHT map in detail. First, we calculated the average Euclidean vowel distance over a regular grid of locations of approximately 1,000 locations by taking the average of all informants who are closest to each grid point. This map is plotted on the left-side map of Figure 25. This controls for the uneven distribution of informants and better shows the overall trend, with most of Connecticut, Rhode Island, and southwestern Massachusetts appearing as areas where the LOT and THOUGHT vowels are generally farther apart. In contrast, most of eastern Massachusetts, Vermont, New Hampshire and Maine appear as areas where the LOT and THOUGHT vowels are generally closer together. Using this map, we then use ordinary kriging (based on a spherical variogram model) to interpolate the LOT/THOUGHT measure across the entire region

(a grid of 300,000 locations), which we plot on the right-side map of Figure 25. This map shows a clear differentiation between the southwest and the rest of New England, which broadly correspond to the underlying patterns visible in the raw map for this variable. Most notably, the map identifies a clear and relatively sharp area of transition between the southwest and the rest of the region, closely aligning with the expected geographic pattern for this variable (see section 5.1).
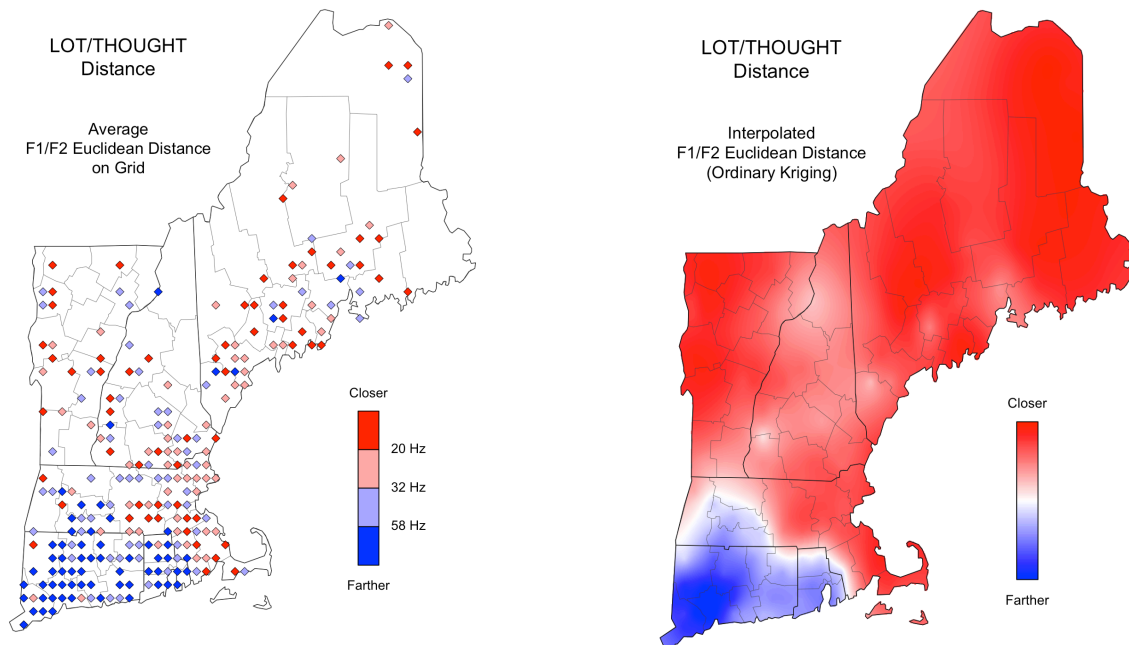

Figure 25. Kriging analysis of LOT/THOUGHT in our audio recordings. Red=closer together. Blue=farther apart. White=in-between.

In other cases, interpolating dialect maps appears to yield somewhat less intuitive results. For example, our interpolated rhoticity map in Figure 26 suggests that r-lessness is, as expected, especially common in southeastern New England and the "down-east" portion of Maine, but it also unexpectedly identifies northeastern Maine as strong region of r-lessness, due to a small number of diverging data points in this relatively rural area of the state. It is important to acknowledge, however, that this is more of an issue with the dataset than with the method for interpolation: the relatively large size of Maine and the limited number of informants in the north makes interpolation less reliable. If we had more data the underlying pattern would be clearer and consequently our analysis would be more reliable. We acknowledge this problematic aspect of applying smoothing methods and consider it an interesting methodological challenge for future work in dialect cartography.[12] Nevertheless, we believe the interpolated maps accurately represent the main patterns of regional variation visible in these five maps, especially in southern New England, from which the majority of our informants originate, and which is for-

---

[12] In addition, a reader reminds us that the kriged rhoticity map depends on binary r-less versus r-ful categorizations of our individual Mechanical Turk speakers (as noted in section 5.2, it was not possible to score each speaker with an r-less percentage since that would require counting each (r) token in 626 speakers' recordings). In other words, our interpolated rhoticity map here would probably be more representative if we had a percentage of r-lessness for each speaker, rather than a binary categorization of each speaker.

tunately where regional differences are strongest. Figure 25 shows LOT/THOUGHT, rhoticity is mapped in Figure 26, Figures 27-28 show the smoothed results for START and PALM,[13] and Figure 29 gives the results for MARY/MARRY/MERRY.

Furthermore, because we have used interpolation to define each of our five acoustic variables across a consistent grid of locations, we are now able to combine these maps to create a single, aggregated map for all the ENE variables analyzed in this study, which we plot in Figure 30. We generated this map simply by scaling and then averaging the interpolated values for all five variables across the 300,000 locations. Although more complex techniques are common in dialectometry (e.g. see Grieve 2016), this relatively simple approach to aggregation was possible because we have a small number of variables. Moreover, the variables clearly show similar east-west divides, especially in southern New England, and we have made our measurements in such a way that in all cases the eastern portions of the map, which always includes Boston, are assigned positive values.
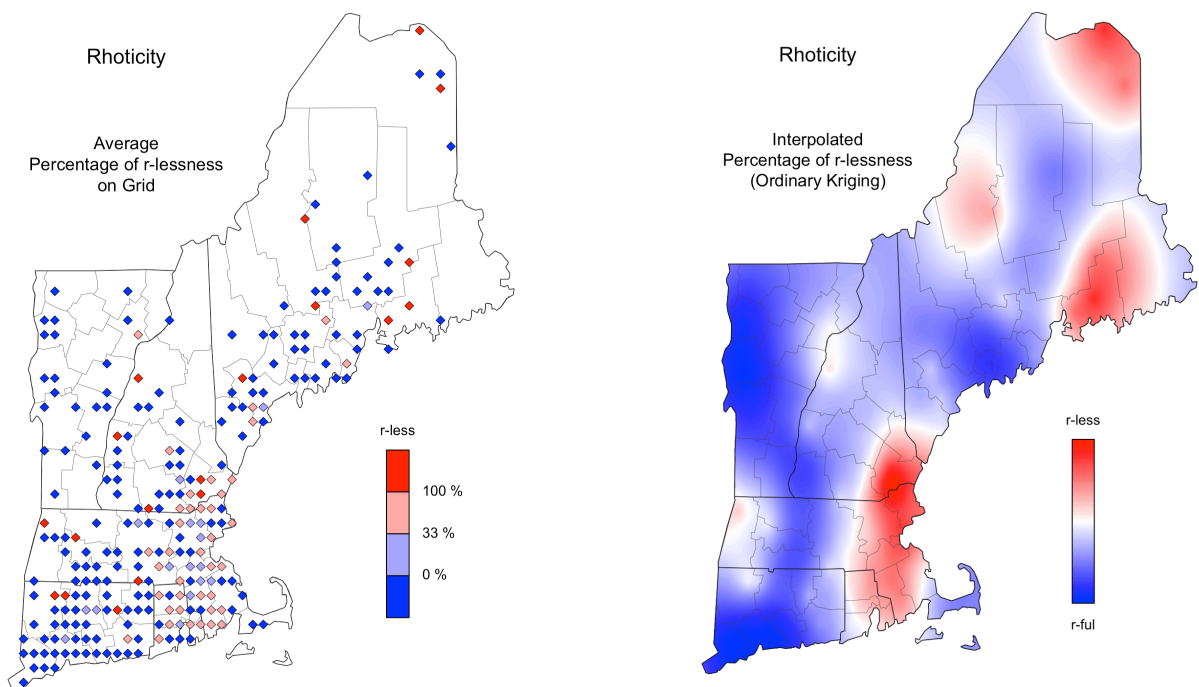
Figure 26. Kriging analysis of the rhoticity data in the audio recordings.

---

[13] Note that the numerical values for PALM- and START-fronting shown here are relative to mean LOT F2 for each given speaker, rather than absolute values. LOT may be subject to variation itself (considering the regional variation in the low-back merger).
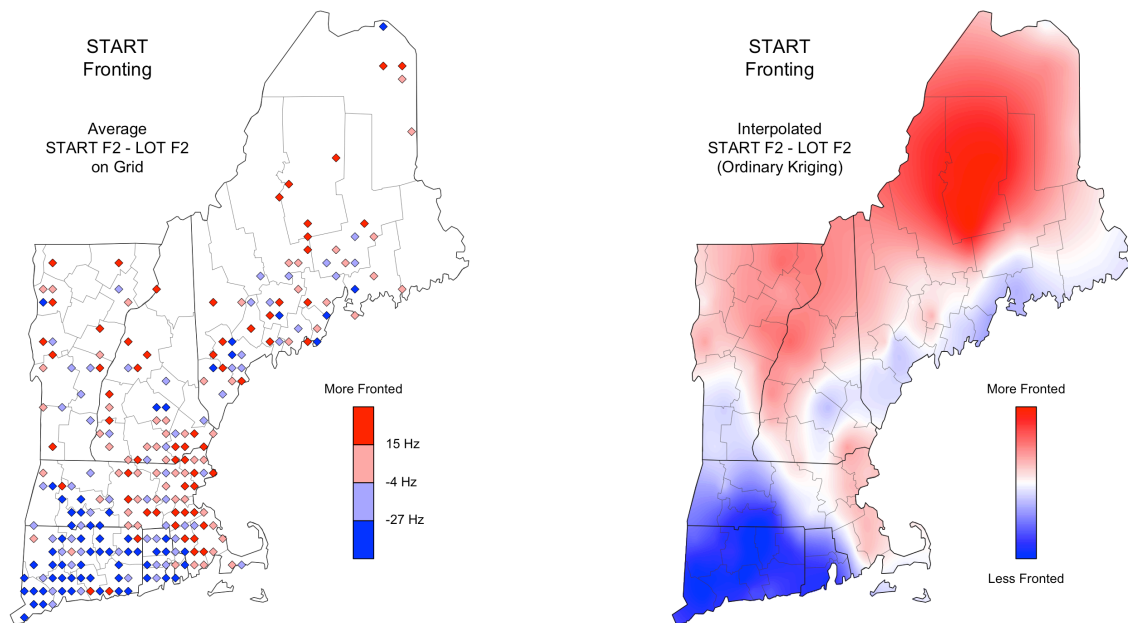
Figure 27. Kriging analysis of START in our audio recordings.
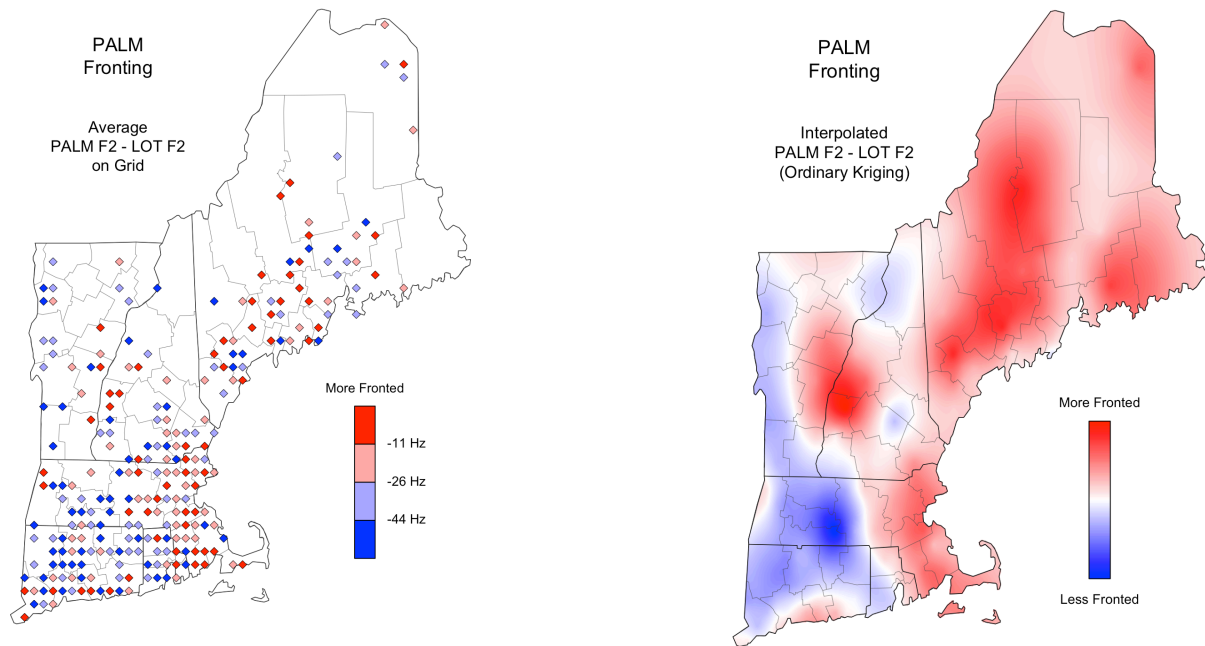
Figure 28. Kriging version of PALM in our audio recordings.
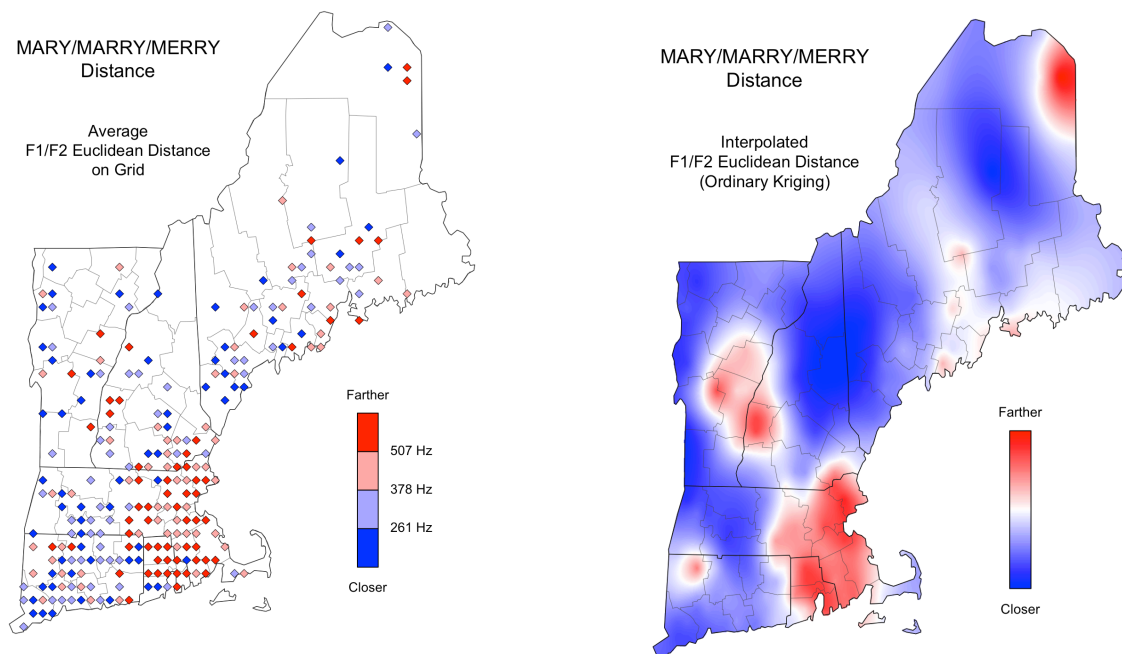
Figure 29. Kriged analysis of MARY/MARRY/MERRY in the audio recordings. Red=closer together, blue=farther apart, white=in-between.
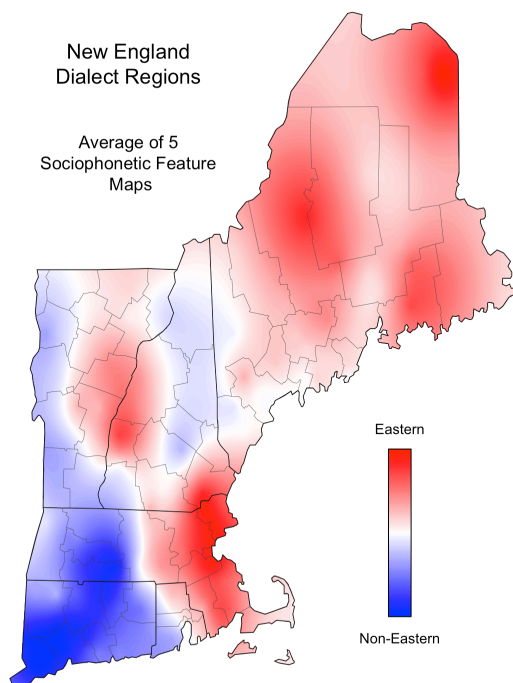
Figure 30. Average interpolated maps for five ENE variables in the study. Red indicates the traditional ENE variants of those variables: r-lessness, fronted PALM, fronted START, MARY/MARRY/MERRY farther apart, and LOT/THOUGHT closer together.

Overall, our aggregated map in Figure 30 identifies a clear distinction between eastern and western New England, with a relatively strong border cutting through Rhode Island, central Massachusetts and southwestern New Hampshire, and western Vermont, near the Green Mountains (see section 2). This pattern closely matches the dialect border between eastern and western New England identified in previous dialect studies, and corresponds closely to historical settlement patterns, as proposed in LANE, with the eastern region primarily settled by colonists originating from eastern Massachusetts, and with the western region primarily settled by colonists originating from southern Connecticut. Indeed, our maps arguably matches these settlement patterns better than the maps in LANE, DARE and ANAE, despite the fact that our dataset is newer than any of these studies, attesting to the value of both our approach to data collection and data analysis.

In addition to replicating previous findings, our aggregated map also suggests some intriguing new possibilities, such as a relatively greater presence of these features in the Upper Valley region of the Connecticut River (NH/VT border area). The aggregate map in Figure 30 (as well as our individual maps in prior sections) hints at Nagy's (2001) notion of a "Live free or die" linguistic separation between parts of New Hampshire and the Boston area. Nagy observed that some New Hampshire residents do not readily adopt Boston residents' phonological patterns despite their proximity to Boston and despite Boston's role as a large metropolis. In Nagy's questionnaire study, some regions of New Hampshire near Massachusetts had fewer ENE features, in contrast with urban eastern Massachusetts. Perhaps this extends to lower Maine as well.

All of our maps, however, should still be interpreted with care. Most important, these maps include all the informants in our dataset, averaging as across all speakers regardless of their demographic background. Although these maps, including this final aggregated map, can therefore be seen as providing a picture of regional variation across the speech community, it is important to stress that factors such as age, gender, social class, ethnicity, and mobility should also be examined before assuming these aggregate geographic patterns apply across social groups. For example, ENE stereotyped features like r-lessness and fronted PALM are known to be rapidly receding in younger generations across northern New England (Stanford et al. 2012, Stanford et al. 2014). Because of how it was constructed, our aggregated map, in particular, can only show a binary distinction between two regions. We have therefore focused on the main division it identifies between eastern and western New England; the additional divisions within eastern New England are interesting, but they should not be over-interpreted. For example, despite apparently being grouped together, the maps should not necessarily be interpreted as evidence that the region along the Maine-New Hampshire border is most similar to western New England.

7. Conclusion

In this online, crowd-sourced Mechanical Turk study of 626 self-recorded speakers and 535 self-reporting questionnaire respondents, we were able to observe many classic geographic patterns of New England variables and uncover new perspectives on the current state of New England dialect features. As the largest acoustic sociophonetic study ever conducted on New England, we believe that the results

provide meaningful progress in the understanding of New England features and American English as a whole. It is evident that New England dialect distinctions can be effectively observed with an online audio-recorded crowdsourcing method and semi-automated acoustic sociophonetic analyses. In addition, such a project makes it possible to build a very large corpus for future research. Most of our Mechanical Turk respondents indicated "yes" when asked if their recordings could be made available to others beyond our research team. We plan to post an accessible version of those recordings in the near future.

As with any research tool, acoustic sociophonetic crowdsourcing of this type has both strengths and weaknesses. Compared to traditional fieldwork, crowdsourcing allows for a much greater geographic range and increased data size in a smaller amount of time and with fewer resources. Crowdsourcing also enables researchers to reach some socioeconomic groups that are not easily accessible from a study done at a university campus. Naturally, online crowdsourcing can never replace the many important benefits of personal, face-to-face fieldwork. Nonetheless, the successful results from this study suggest that a crowdsourcing method like Amazon Mechanical Turk can be a valuable, complementary approach to traditional methods, and we hope that future studies will use similar crowdsourcing tools in other geographic regions. Bring on the crowd!

Acknowledgements

Appendix A: Written Materials in the Mechanical Turk Audio Recording Project

The following is a copy of the demographic information and the audio activities of the Mechanical Turk audio recording project.

To participate in this survey, you must be over 18 years old and have spent most of your childhood in one of these states: Vermont, New Hampshire, Maine, Connecticut, Rhode Island, or Massachusetts.

Please answer a few questions about yourself and then record yourself reading a series of short passages out loud. You must complete *all* of the questions in the survey.

* denotes questions with required responses.
What is your gender?*
Male Female Other

In which year were you born?* Enter in YYYY (4-digit) format.

Which of the following US Census categories best represents your ethnicity?*
[US census categories listed in a drop-down menu]

During ages 0-12, in which New England state did you spend the most time?*
[Six New England states listed in a drop-down menu]

During ages 0-12, what is the name of the city/town you spent the most time in?*
Please list the ONE city/town which best answers this question. If it's the Boston area, please give the specific location, such as "Charlestown", "Mattapan", "Lexington", etc.

During ages 0-12, what is the 5 digit zip code in which you spent the most time? Optional; leave blank if unknown.

During ages 0-12, which of the following best describes your location?*
      rural suburban urban

During ages 13-18, in which New England state did you spend the most time?*
      [Six New England states listed in a drop-down menu]

During ages 13-18, what is the name of the city/town you spent the most time in?*
Please list the ONE city/town which best answers this question. If it's the Boston area, please give the specific location, such as "Charlestown", "Mattapan", "Lexington", etc.

During ages 13-18, what is the 5 digit zip code in which you spent the most time? Optional; leave blank if unknown.

During ages 13-18, which of the following best describes your location?*
      rural suburban urban

After age 18, in which US State (or DC) did you spend the most time?*
      [US states listed in a drop-down menu, also "Not in the US"]

After age 18, what is the name of the city/town you spent the most time in?*
Please list the ONE city/town which best answers this question. If it's the Boston area, please give the specific location, such as "Charlestown", "Mattapan", "Lexington", etc.

After age 18, what is the 5 digit zip code in which you spent the most time? Leave blank if unknown.

Which of the following best describes your highest achieved education level?*
      Drop-down menu: Some high school or less   High school
                            Some college, no degree   Associate degree
                            Bachelor's degree   Graduate degree

Please enter your occupation. If currently unemployed, please enter your most recent occupation.*
If you are a student, enter the occupation of the primary income source in your household when growing up.

[After a consent process (not shown here), participants moved to a page with the recording task]

Recording Task (part 1 of 12)

How to record audio on your computer

Click for the instructions on recording audio in your operating system. You may also use any other audio recording program that you a

Mac Users

Windows Users

Record and Upload

Read aloud and record the passage below (1 of 12). Each sentence should be read twice as shown.

Please use your normal everyday voice, not deliberately polished or formal. Then upload the recorded file. Any filename is fine -- just i upload the file that contains your recording of the sentences on this current page.

> I hope that Mary bought coffee and pizza at the food shop.
> I hope that Mary bought coffee and pizza at the food shop. *

Choose File   no file selected
Follow the instructions to record yourself reading this passage and save the file. The passage should be read TWICE as shown.

submit

[See section 3 for a list of all twelve sentences that the participants read.]

Appendix B: The self-reporting questionnaire

The following is a copy of the demographic information and the survey activities of the self-reporting lexical/phonological questionnaire project:

For the questions about dialect features below, we used Vaux and Golder's survey (2003) as a model, sometimes using the same wording.

To participate in this survey, you must be over *18 years old* and have spent *most of your childhood in these areas*: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont.

Birth year (4 digits): _____

Birth place:

_____ _____ _____
city/town   state   zipcode (if known)
Where did you live during the majority of the your *childhood* years?
Please give one location that best answers this question:

_____ _____ _____
city/town   state   zipcode (if known)

Where did you live during the majority of the your *teenage* years?
Please give one location that best answers this question:

_____ _____ _____
city/town   state   zipcode (if known)

Where have you lived during the majority of the your *adult* years so far?
Please give one location that best answers this question:

_____ _____ _____
city/town   state   zipcode (if known)

Gender: female   male   other: _____

Ethnicity: Which of the following US census categories best represents your ethnicity? Please select just <u>one</u>:
White, Black or African American, Native American/American Indian/Alaska Native, Chinese, Japanese, Filipino, Korean, Asian Indian, Vietnamese, Other Asian, Native Hawaiian, Samoan, Guamanian or Chamorro, Other Pacific Islanders, Mexican/Mexican American/Chicano, Puerto Rican, Cuban, Another Hispanic/Latino/Spanish origin, Other:_____

Which of the following best describes your highest achieved education level?
Some high school, Some college but no degree, Associates degree, Bachelors degree, Graduate degree (Masters, Doctorate, etc)

Please list your occupation. *If currently unemployed, please list your most recent occupation.
*If you are a student, please list the occupation of the primary income source in your household when growing up:
_____

For each of the following questions, please choose the one best answer. We realized that you may have more than one word for some of these items, so please just choose the word you use the most. [Note: some questions are based on Vaux & Golder's (2003) survey and Nagy (2001)]

[Note: In the following, we have omitted all questions about lexical items since those results are not discussed in the present paper.]

5. When you say *father* and *bother*, do they rhyme (like *feather* and *weather*)? [wording based on Nagy 2001]

a. *Father* and *bother* rhyme in my pronunciation   b. *Father* and *bother* do not rhyme in my pronunciation
c. Not sure

6. Do you pronounce the words *cot* and *caught* the same or different?
a. Same b. Different   c. Not sure

7. Do you pronounce the three words *Mary, merry, marry* the same or differently?
a. I pronounce all three the same
b. I pronounce all three differently
c. *Mary* and *merry* are the same, but *marry* is different
d. *Mary* and *marry* are the same, but *merry* is different
e. *merry* and *marry* are the same, but *Mary* is different

10. Do you "drop the r" in words like *card* and *weather*?
a. Yes   b. No   c. Sometimes

11. Do you personally <u>know people</u> who "drop the r" in words like *card* and *weather*?
a. Yes   c. No, but I sometimes hear it   d. No, and I never hear it

46. Are there any other distinctive words or pronunciations from your region that you'd like to briefly tell us about? [open format answer]

References

Ash, Sharon (2013). Social class. In J. K. Chambers & Natalie Schilling-Estes (eds.), *The handbook of language variation and change,* 2nd ed. Malden: Blackwell. 327-67.

Atkinson, Peter, & Christopher Lloyd (2009). Geostatistics and spatial interpolation. In A. Stewart Fotheringham & Peter Rogerson (eds.), *The Sage handbook of spatial analysis.* London: Sage. 159-182.

Bivand, Roger, Edzer Pebesma & Virgilio Gomez-Rubio (2008). *Applied spatial data analysis with R.* New York: Springer.

Boberg, Charles (2001). The phonological status of western New England. *American Speech* 76:3–29.

Bowern, Claire (2010). North American English dialect survey. Yale University.

Britain, David, A. Leemann, and M. J. Kolly (2016). Up, app and away?: Social dialectology and the use of smartphone technology as a data collection strategy. In *Sociolinguistics Symposium* 21. Murcia, Spain: University of Murcia.

Browne, Charlene & James Stanford (forthcoming). Boston dialect features in the Black/African American community. *Penn Working Papers in Linguistics* 24.2, Selected Papers from NWAV 46.

Burnett, Heather (2012). The role of microvariation in the study of semantic universals: Adverbial quantifiers in European and Québec French. *Journal of Semantics* 29:1-38.

Carver, Craig (1987). American regional dialects. Ann Arbor: University of Michigan.

Cassidy, Frederick (ed.) (1985). *Dictionary of American Regional English*, volume 1. Cambridge: Harvard University Press/Belknap.

Chartier, Nicole (2013). My father doesn't bother with the Farm: The changing state of the low-back vowel in southeastern New Hampshire. Ms. University of New Hampshire.

DARE. *Dictionary of American Regional English* (1985-2013). Edited by Frederic G. Cassidy and Joan Houston Hall. 6 vols. Cambridge, Mass.: Belknap Press of Harvard University Press.

Dinkin, Aaron J. (2005). Mary, darling, make me merry; say you'll marry me: Tense-lax neutralization in the Linguistic Atlas of New England. In *University of Pennsylvania Working Papers in Linguistics* 11(2):73-90.

Dinkin, Aaron (2009). Dialect boundaries and phonological change in upstate New York. PhD dissertation, University of Pennsylvania.

Evanini, Keelan (2009). The permeability of dialect boundaries: A case study of the region surrounding Erie, Pennsylvania. Ph.D. dissertation, University of Pennsylvania.

Evanini, Keelan, Stephen Isard & Mark Liberman (2009). Automatic formant extraction for sociolinguistic analysis of large corpora. *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (INTERSPEECH), September 6–10, Brighton, UK. 1655-58.

Fruehwald, Josef (2014). Automation and sociophonetics. Presented at *New Ways of Anlayzing Variation 43*, October 23, Chicago.

Gorman, Kyle, Jonathan Howell & Michael Wagner (2011). Prosodylab-Aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics* 39.3. 192–193.

Grieve, Jack (2015). A statistical comparison of regional phonetic and lexical variation in American English. *Literary and Linguistic Computing* 28: 82-107.

Grieve, Jack (2017). Spatial statistics for dialectology. In Charles Boberg, John Nerbonne, & Dominic Watt (eds.) *The Handbook of Dialectology.* Wiley Blackwell.

Heselwood, Barry, Leendert Plug & Alison Tickle (2008). Assessing rhoticity using auditory, acoustic and psychoacoustic methods. In Clive Upton & Barry Heselwood (eds) *Proceedings of Methods XIII: Papers from the Thirteenth International Conference on Methods in Dialectology*. 331-40.

Johnson, Daniel Ezra (2009). Getting off the GoldVarb standard: Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1):359-83.

Johnson, Daniel Ezra (2010). *Stability and change along a dialect boundary: The low vowels of southeastern New England*. PADS No. 95. Durham, North Carolina: Duke University.

Johnson, Daniel Ezra, & David Durian (2017). New England. In Raymond Hickey (ed.), *Listening to the past: Audio records of accents of English*. Cambridge: Cambridge University Press. 257-97.

Kendall, Tyler, & Valerie Fridland (2017). Regional relationships among the low vowels of U.S. English: Evidence from production and perception. *Language Variation and Change* 20(2):245-71.

Kendall, Tyler, & Erik Thomas (2010). Vowels: vowel manipulation, normalization, and plotting in R. R package. cran.r-project.org/web/packages/vowels/index.html [computer software]

Krige, D.G. (1951). *A statistical approach to some mine valuations and allied problems at the Witwatersrand*. Master's thesis, University of Witwatersrand.

Kurath, Hans (1939). *Handbook of the linguistic geography of New England*. Providence: Brown.

Kurath, Hans (1949). *Word geography of the eastern United States*. Ann Arbor: University of Michigan.

Kurath, Hans, & Raven McDavid, Jr. (1961). *The pronunciation of English in the Atlantic States*. Ann Arbor: University of Michigan Press.

Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (2001). *Principles of linguistic change, vol. 2, social factors*. Oxford: Blackwell.

Labov, William, Sharon Ash & Charles Boberg (2006). *The Atlas of North American English*. Berlin: Mouton.

Labov,William, Ingrid Rosenfelder & Josef Fruehwald (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal and reanalysis. *Language* 89(1):30–65.

Laferriere, Martha (1979). Ethnicity in phonological variation and change. *Language* 55:603-17.

LANE. *Linguistic Atlas of New England* (1939-43). Edited by Hans Kurath, with the collaboration of Miles L. Hanley, Bernard Bloch, Guy S. Lowman, Jr., and Marcus L. Hansen. 3 vols. Providence, R.I.: Brown University.

Leemann, A., M. J. Kolly, R. Purves, D. Britain & E. Glaser (2016). Crowdsourcing language change with smartphone applications. *PLoS One*, 11(1): 1-25.

Mufwene, Salikoko (1996). The Founder Principle in Creole genesis. *Diachronica* 13.1: 83-134.

Mufwene, Salikoko (1999). The Founder Principle revisited: Rethinking feature selection in North American English. Paper presented at the *Tenth International Conference on Methods in Dialectology*, St. John's, Newfoundland, Aug. 1–6.

Nagy, Naomi (2001). "Live Free or Die" as a linguistic principle. *American Speech* 76.1:30-41.

Nagy, Naomi & Patricia Irwin (2010). Boston (r): Neighbo(r)s nea(r) and fa(r). *Language Variation and Change* 22:241-78.

Nagy, Naomi & Julie Roberts (2004). New England: Phonology. In Edgar Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds), *Handbook of varieties of English: The Americas and the Caribbean*. Berlin/New York: Mouton de Gruyter.

Purnell, Thomas (2012). Dialect recordings from the Hanley Collection, 1931-1937. *American Speech* 87(4):511-13.

Ravindranath, Maya (2011). A wicked good reason to study intensifiers in New Hampshire. *New Ways of Analyzing Variation 40*, Georgetown University, October 27-30.

Reddy, Sravana & James Stanford (2015). Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard.* DARLA is available at darla.dartmouth.edu

Roberts, Julie (2006). As old becomes new: Glottalization in Vermont. *American Speech* 81(3):227-49.

Roberts, Julie (2007). Vermont lowering? Raising some questions about /ai/ and /au/ south of the Canadian border. *Language Variation and Change* 19:181-197.

Roberts, Julie (2016). Internal boundaries and individual differences: /aʊ/-raising in Vermont. *American Speech* 91(1):34-61.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, Jiahong Yuan (2014). FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2. [computer software]

Sankoff, Gillian (2004). Adolescents, young adults and the critical period: Two case studies from 7-Up. In Carmen Fought (ed) *Sociolinguistic variation: Critical reflections*. Oxford: Oxford University Press. 121-39.

Sankoff, Gillian and Hélène Blondeau (2007). Language change across the lifespan: /r/ in Montreal French. *Language* 83(3):560-88.

Severance, Nathan, Keelan Evanini & Aaron Dinkin (2016). Examining the reliability of automated vowel analysis using FAVE. Paper presented at *NorthWest Phonetics & Phonology Conference* (NoWPhon), Eugene, Oregon. May 13-15.

Sipple, Kelsey, James Stanford, & Dartmouth LING 80 class (2015). Boston Strong: South Boston dialect features across 70 years of apparent time. Paper presented at *Linguistic Society of America Annual Meeting*, January 9, Portland, Oregon.

Stanford, James N., Thomas Leddy-Cecere & Kenneth Baclawski (2012). Farewell to the founders: Major dialect changes along the east-west New England border. *American Speech* 87(2):126-69.

Stanford, James N., Nathan Severance & Kenneth Baclawski, Jr. (2014). Multiple vectors of unidirectional dialect change in Eastern New England. *Language Variation and Change* 26(1):103-140.

Stanley, Joey, & Margaret Renwick (2016). Phonetic shift /ɔr/ phonemic change? American English mergers over 40 years. *Laboratory Phonology* (LabPhon15).

Thomas, Erik (2001). *An acoustic analysis of vowel variation in New World English*. Publication of the American Dialect Society 85. Durham, NC: Duke University Press.

Thomas, Erik (2010). A longitudinal analysis of the durability of the northern/midland dialect boundary in Ohio. *American Speech* 85(4):375-430.

Vaughn, Charlotte & Tyler Kendall (2016). What do listeners know about about the sociolinguistic variable (ING)? *New Ways of Analyzing Variation* (NWAV-45). Vancouver, Canada. November.

Vaux, Bert & Scott Golder (2003). The Harvard Dialect Survey. Cambridge, MA.

Villard, Sarah (2009). Postvocalic /r/ in the Upper Valley of Vermont and New Hampshire. Masters thesis, University of New Hampshire.

Wagner, Suzanne Evans (2012). Age grading in sociolinguistic theory. *Language and Linguistics* 6(6): 371-382.

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Wolfram, Walt & Natalie Schilling (2016). *American English.* Malden, MA: Wiley Blackwell.

Wood, Jim (2011). Short-a in northern New England. *Journal of English Linguistics* 39(2):135-65.

Wood, Jim, Laurence Horn, Raffaella Zanuttini & Luke Lindemann (2015). The Southern Dative Presentative meets Mechanical Turk. *American Speech* 90:291-320.

Yaeger-Dror, Malcah, Tyler Kendall, Paul Foulkes, Dominic Watt, Phillip Harrison, Daniel Johnson, Colleen Kavanagh & Jillian Oddie (2009). Perception of 'r' by trained listeners. Presented at *Linguistic Society of America Annual Meeting,* January 8-11, San Francisco.

Zelinsky, Wilbur (1992). *The cultural geography of the United States*. Rev. ed. Englewood Cliffs, N.J.: Prentice Hall.