



Published in final edited form as:

Speech Commun. 2006 June 1; 48(6): 633–644. doi:10.1016/j.specom.2005.09.010.

The Nationwide Speech Project: A new corpus of American English dialects[★]

Cynthia G. Clopper^{a, *} and David B. Pisoni^b

^a Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA

^b Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

Abstract

Perceptual and acoustic research on dialect variation in the United States requires an appropriate corpus of spoken language materials. Existing speech corpora that include dialect variation are limited by poor recording quality, small numbers of talkers, and/or small samples of speech from each talker. The Nationwide Speech Project corpus was designed to contain a large amount of speech produced by male and female talkers representing the primary regional varieties of American English. Five male and five female talkers from each of six dialect regions in the United States were recorded reading words, sentences, passages, and in interviews with an experimenter, using high quality digital recording equipment in a sound-attenuated booth. The resulting corpus contains nearly an hour of speech from each of the 60 talkers that can be used in future research on the perception and production of dialect variation.

Keywords

Speech corpus; Dialect variation; American English

1. Introduction

1.1. Phonological dialect variation in the United States

Researchers have been documenting regional linguistic variation in the United States for more than a century. The American Dialect Society was founded in 1889 with the goal of collecting a comprehensive American English dictionary. Krapp (1925) documented regional varieties of American English based on grammar and pronunciation guides dating back to the 18th century. From this research, he identified three main dialects of American English: Eastern, Southern, and Western (or General American). Thirty years later, McDavid (1958) described the early Linguistic Atlas projects in the United States, which documented lexical and phonological variation based on fieldwork interviews conducted in predominantly rural areas. McDavid (1958) concluded that the major dialects of American English were Northern, Midland, and Southern. He also acknowledged that these dialects were more sharply distinguished on the Atlantic seaboard and that more transition areas between dialects were found as one moved westward across the country. Carver (1987) also

[★]Parts of this research were presented at the 147th Meeting of the Acoustical Society of America in New York, NY, May 24–28, 2004.

^{*}Corresponding author. c-clopper@northwestern.edu (C.G. Clopper), pisoni@indiana.edu (D.B. Pisoni).

examined regional lexical variation but he described only two primary dialects of American English: Northern and Southern.

More recently, linguistic variation has been explored using acoustic–phonetic analysis techniques. Thomas (2001) obtained acoustic–phonetic vowel spaces for nearly 200 individual talkers. Although Thomas (2001) made no explicit claims about specific dialect regions, he did group his talkers into a Northern group and a Southern group. Labov and his colleagues (forthcoming) have been working on a more comprehensive study of regional variation in American English. The Telephone Survey (TELSUR) project at the University of Pennsylvania includes telephone interviews with 700 talkers representing all major urban areas in the United States. The recordings have been analyzed acoustically and Labov (1998) defined three major dialects of American English based on the vowel systems of his 700 talkers: Northern, Southern, and the “third dialect”. This third dialect includes Eastern and Western New England, Western Pennsylvania (centered on Pittsburgh), the Midland, and the West. Labov (1998) described the Mid-Atlantic metropolitan areas from New York City to Washington, DC as “exceptions” to the three-dialect division because speakers from this region do not exhibit the characteristic properties of any of the three major dialects. The Florida peninsula is also treated as a unique region because of the high level of dialect mixing that occurs there as a result of migration from other states. Fig. 1 is a map of the United States showing these regions, based on the work by Labov and his colleagues (forthcoming). No data are available for the gray areas on the map because these regions are sparsely populated and the TELSUR project focused on 145 urban areas with an average population of 1.7 million people, ranging from 88,000 in Aberdeen, South Dakota to 17.6 million in New York City (Ash, n.d.).

The vowel system of the Northern dialect of American English is characterized by the Northern Cities Chain Shift (Labov, 1998). The Northern Cities Chain Shift is a clockwise shift of the low vowels that includes the fronting and raising of /æ/, the fronting of /ɑ/, the lowering of /ɔ/, and the backing of /ʌ/ and /ɛ/. /ɪ/ is also reported to be backed in the Northern dialect as a parallel shift to /ɛ/ backing. Fig. 2 depicts the major features of the Northern Cities Chain Shift.

The Southern dialect of American English is characterized by the Southern Vowel Shift (Labov, 1998). The primary feature of this shift is the fronting of the back vowels /u/ and /o/. In addition, the front lax vowels /ɪ/ and /ɛ/ are raised in Southern American English and the front tense vowels /i/ and /e/ are lowered. The Southern Vowel Shift is shown in Fig. 3. The Southern dialect is also characterized by the monophthongization of the diphthongs /ɑy/ and /oy/ (Thomas, 2001).

The common feature of the “third dialect” of American English is the merger of the low back vowels /ɑ/ and /ɔ/, creating homophones of such pairs of words as *caught* and *cot* or *Dawn* and *Don* (Labov, 1998). The subdialects of the “third dialect” also have some unique features of their own. Other features of Eastern New England include raising of the nucleus in the diphthongs /ɑy/ and /aw/ (Thomas, 2001). Western New England, on the other hand, reflects components of the Northern Cities Chain Shift with some raising of /æ/, fronting of /ɑ/, and backing of /ɛ/ (Boberg, 2001; Thomas, 2001). Western speech is characterized by the low-back merger and by /u/ fronting (Labov et al., forthcoming; Thomas, 2001). Unlike back vowel fronting in the South, which has spread from /u/ to /o/, however, the Western pattern is typically limited to fronting of /u/. The Midland dialect is the least marked of the regional American English varieties, exhibiting no distinct features other than the “third dialect” /ɑ/ ~ /ɔ/ merger.

As mentioned above, the Mid-Atlantic dialect does not exhibit the “third dialect” /ɑ/ ~ /ɔ/ merger and in fact, the two vowels are more distinct due to /ɔ/ raising (Labov, 1994; Thomas, 2001). The maintenance of a historical contrast between long and short /æ/ in the Mid-Atlantic region means that /æ/ is also raised in some words, but not others (Labov, 1994; Thomas, 2001).

1.2. Considerations in designing new corpora

A number of factors must be considered when designing and collecting a speech corpus, including the demographics of the talkers and the interviewer(s), the recording equipment and conditions, and the types of speech materials to be collected. Table 1 provides examples of each of these factors. Talker demographics include age, gender, socioeconomic status, race, ethnicity, level of education, residential history, and linguistic experience. Residential history can include the region of origin of the talker as well as the number of different places he or she has lived. Linguistic experience includes the talkers’ native language and any foreign language experience and exposure. The experimenter must decide whether or not to control for each of these variables. The decision to include specific variables or exclude others is related to the ultimate goals of the corpus. For example, if the primary use of the corpus will be comparisons between certain linguistic forms across gender, the experimenter would want to design a corpus that is balanced for gender. If, however, the experimenter is interested only in the speech of female newscasters, the corpus could be limited to female talkers.

Research in social psychology has shown that talkers often accommodate their speech to that of their interlocutor (Giles and Powesland, 1997; Trudgill, 1998). Therefore, given the potential for stylistic shifts in a talker’s speech due to perceived social differences between the interviewer and the talker, the demographics of the interviewer(s) must also be considered. In addition to considering the same demographic factors described above for the talkers, the experimenter must also determine whether the interviewers should be “insiders” or “outsiders” to the community or communities that the speech corpus represents (Feagin, 2002; Wolfram and Fasold, 1997). As a result of these issues, some corpora rely on a number of different interviewers with different backgrounds, while others rely on only a single interviewer.

The experimenter must also balance two aspects of the recording equipment and conditions: quality of the recordings and speaking style. While sound-attenuated booths and high quality digital recording equipment lead to high quality recordings, such formal settings also typically encourage the production of so-called “lab speech” (Labov, 1972b; Rischel, 1992). On the other hand, fieldwork practices typically result in more natural, conversational speech, but these samples are obtained using poorer quality recording devices or have more substantial background noise (Plichta and Mendoza-Denton, 2001). In order to determine the most appropriate recording conditions for a given corpus, the goals of the project must be considered. For example, if the recordings are to be used to document lexical and phonological variation, field recordings may be acceptable. However, if the recordings are to be used in acoustic analyses or playback experiments with naïve listeners, higher quality recordings would be preferred.

Finally, the experimenter must decide what kinds of speech materials to collect. Traditional sociolinguistic research is based on interview speech in which the informants respond to questions designed to elicit specific lexical items or are asked open-ended questions about childhood games, near-death experiences, or the local community (Labov, 1972a). Most speech perception and spoken word recognition research, on the other hand, is based on read speech produced in the laboratory in order to control for the lexical, segmental, and prosodic content of the utterances. Numerous studies have shown that speaking style (e.g., read

speech vs. conversational speech) affects the degree to which certain regional or ethnic dialect variables are produced, with fewer stigmatized forms appearing in read speech than in interview speech (Labov, 1972a). Several recent speech corpora contain some samples of both read and spontaneous speech in order to provide both “natural” and linguistically controlled utterances (e.g., the TELSUR Project; Labov et al., forthcoming).

1.3. Existing spoken language corpora with dialect variation

A number of corpora currently exist that contain variation due to the regional and ethnic background of the talkers. A summary of the features of these five corpora is shown in Table 2. Because these corpora were collected with different goals and intended uses, they all have some strengths and some weaknesses related to the factors described above. For example, the Dictionary of American Regional English (DARE) project includes field-work recordings of interviews with individuals in more than 1000 communities across the United States collected between 1965 and 1970 (Hall and von Schneidmeyer, 2004). The interviews included more than 1800 questions and also included a reading of the *Arthur the Rat* passage, a short narrative designed to elicit regional phonological variation. The talkers in the DARE interviews differed in terms of age, ethnicity, socioeconomic status, and gender. The recordings have been used primarily in the production of the multi-volume Dictionary of American Regional English (1985) which describes lexical variation in the United States in detail. The strengths of this corpus include the large number of talkers and the large samples of speech from each talker. The weaknesses include its poor recording quality and its uneven distribution of talker demographics.

Another example of a speech corpus that includes traditional sociolinguistic interviews is the recent SLX Corpus of Classic Sociolinguistic Interviews (Strassel et al., 2003). This corpus includes eight interviews with a total of nine different talkers producing a range of utterances including narratives, interview responses, and word lists. The primary strength of the SLX corpus is its utility as a pedagogical tool for training sociolinguistic field-workers. As with the DARE interviews, the recording quality of the materials is relatively poor, although the SLX corpus contains digital files recorded from the original fieldwork tapes. Another weakness of the SLX corpus is the relatively small number of talkers included.

The Santa Barbara Corpus of Spoken American English (DuBois et al., 2000) is another source for speech samples containing regional and ethnic variation. The Santa Barbara corpus includes hundreds of recordings of “natural” speech including conversations, political speeches, classroom lectures, and bedtime stories. The talkers differ in terms of age, gender, ethnicity, region of origin, and socioeconomic status. The materials from the Santa Barbara corpus are particularly well suited for studies of prosodic variation and discourse analysis. The main strengths of the corpus are its wide range of speaking styles and speech materials. Its weaknesses include variable recording conditions and an uneven distribution of talker demographics.

One set of corpora that explicitly matched talkers in terms of regional dialect is the CallFriend project (Canavan and Zepperlen, 1996a,b). One of these corpora includes 60 telephone conversations between two speakers of Southern American English and the other includes 60 telephone conversations between two speakers of non-Southern varieties of American English. Together these corpora provide an excellent source for materials for spoken word recognition research. The main strength of this corpus is the large number of talkers who were recorded. Its weaknesses include the limited bandwidth of the telephone recordings and the fact that the assignment of the talkers to the Southern or non-Southern corpus was based on samples of speech from each talker, not his or her true residential history.

The only existing corpus of regional variation in the United States that obtained high quality audio recordings in a sound-attenuated booth is the TIMIT Acoustic–Phonetic Continuous Speech Corpus (Fisher et al., 1986; Zue et al., 1990). The TIMIT corpus contains recordings of 630 talkers who each read 10 different sentences. Age, gender, ethnicity, level of education, height, and regional dialect are provided for each talker. The TIMIT corpus was originally designed for use in automatic speech recognition research, although it has also been used recently in acoustic analyses of regional and gender-based variation (Byrd, 1994; Clopper and Pisoni, 2004b), perceptual dialect categorization experiments (Clopper et al., 2005a; Clopper and Pisoni, 2004a,b), and automatic dialect classification (Rojas, 2002). The strengths of the TIMIT corpus include the high quality of the audio recordings and the large number of talkers. The main weakness of the TIMIT is the limited amount of speech from each talker. In addition, the regional labels assigned to the talkers do not accurately reflect the major regional varieties of American English that Labov et al. (forthcoming) have proposed and it is unclear what criteria were used to assign the regional labels to the talkers.

Each of the five corpora described above was designed for a different purpose and the strengths and weaknesses of each corpus reflect those varied goals. A corpus such as the DARE recordings covers a large amount of geographic and lexical territory, but is limited by the quality of the recordings. The TIMIT corpus, on the other hand, provides high quality recordings for a large number of talkers, but the speech materials produced by each talker are severely limited.

2. New corpus design and collection

The Nationwide Speech Project (NSP) corpus was designed to provide a large amount of speech from male and female talkers representing a number of different regional varieties of American English for use in acoustic analyses and perceptual tasks with naïve listeners. Nearly an hour of speech was collected from each talker producing a range of speaking styles from isolated read words to interview speech. The read words, sentences, and passages can be used in acoustic analyses and playback experiments when it is desirable to have identical linguistic content across all of the talkers. The interview speech samples can be used for projects in which more “natural” or continuous speech samples are desired. To allow for precise acoustic measurements and to reduce the effects of non-linguistic artifacts in the recorded stimulus materials, the recordings were made using high quality digital equipment in a sound-attenuated booth. Finally, the demographic variables of the talkers were strictly controlled such that the resulting corpus includes speech samples from a relatively homogeneous population of talkers that vary only by gender and region of origin.

The NSP corpus has several features which distinguish it from the TIMIT corpus and the TELSUR project. First, the NSP corpus contains more speech samples per talker than either the TIMIT or TELSUR projects, although the total number of talkers is substantially less and the geographic distribution of the talkers is more limited in the NSP corpus. Second, unlike the TIMIT corpus, the NSP corpus was designed using sociolinguistically motivated regions and each of the selected regions was represented by the same number of talkers. In this way, the NSP corpus is a better reflection of linguistic variation than the TIMIT corpus. Third, unlike the TELSUR project, the recordings for the NSP corpus were made using high quality digital recording equipment to allow for more accurate acoustic–phonetic analyses, as well as perceptual experiments. Thus, the NSP corpus can be used in speech science experiments as well as sociolinguistic or sociophonetic research.

2.1. Stimulus materials

Four different kinds of speech materials were collected from each talker in the NSP corpus: isolated words, sentences, passages of connected speech, and interview speech. Table 3

shows examples of the materials collected for the NSP corpus. The isolated words were divided into three sets of materials: hVd words, CVC words, and multi-syllabic words. The hVd words consisted of five repetitions of each of 10 American English vowels in the hVd context: *heed, hid, hayed, head, had, hod, hud, hoed, hood*, and *who'd*. The CVC wordlist was composed of 76 monosyllabic English words. Each of the 14 monophthongal and diphthongal vowels in American English was included at least four times in the CVC list and the following consonantal context for each vowel was varied to include liquids, nasals, and voiceless and voiced obstruents. The multi-syllabic word list was a subset of 112 of the stimulus materials originally designed by Carter and Clopper (2002) for their study of word reduction by normal-hearing adults. The words in the list were balanced for number of syllables (two, three, or four), location of primary stress (first, second, or third syllable), and morphological complexity (monomorphemic or polymorphemic). All of the words in the CVC and multi-syllabic lists were highly familiar and received a familiarity rating of at least 6.0 (on a seven-point scale) by Indiana University undergraduates (Nusbaum et al., 1984).

The read sentence materials were also divided into three sets of materials: high probability sentences, low probability sentences, and semantically anomalous sentences. The high probability and low probability sentences were taken from the Speech Perception in Noise (SPIN) test (Kalikow et al., 1977). Examples of the high probability and low probability sentences are shown in Table 3. The SPIN sentences range in length from five to eight words and are phonetically balanced with respect to phoneme frequency in English. High probability sentences are defined as having a final target word that is predictable from the preceding semantic content of the sentence. Low probability sentences have a final target word that is not predictable from the preceding sentence context. The low probability sentences in the SPIN test were created by placing each high probability target word at the end of one of several generic sentence contexts such as, "I did not know about the ...". Thus, in the original SPIN test, all of the high probability sentences were paired with a low probability sentence with the same target word (see Kalikow et al., 1977). For the NSP corpus, 102 high probability sentences and 52 low probability sentences were selected from the SPIN test. The low probability sentences were each paired with a high probability sentence with the same target word. The remaining 50 high probability sentences were not paired with a low probability sentence.

The anomalous sentence list was created specifically for the NSP corpus. Using the high probability sentences as a syntactic frame, each content word was replaced with a different content word from the same syntactic class (e.g., noun, verb, or adjective). The target word in each sentence was left unchanged. The resulting utterances were semantically anomalous but syntactically correct sentences. Examples of the anomalous sentences are shown in Table 3. Each anomalous sentence was structurally parallel to one of the high probability sentences. In addition, the target words of the 52 anomalous sentences were matched with the target words of the 52 low probability sentences and the corresponding 52 high probability sentences. To ensure that the anomalous sentences were roughly equivalent in their semantic anomaly, all of the sentences were presented visually to a group of naïve participants who were asked to rate them on a seven-point sensible/strangeness scale. Sentences rated more than one standard deviation above or below the mean were revised (see Clopper et al., 2001).

In addition to the three word lists and the three sentence lists, each talker also read two meaningful passages: the first paragraph of the *Rainbow Passage* (Fairbanks, 1940) and the entire *Goldilocks Passage* (Stockwell, 2002). The *Rainbow Passage* has been used in a variety of acoustic and perceptual studies of speech, including investigations of talker differences (e.g., Gelfer and Schofield, 2000) and the speech of clinical populations (e.g., Baker et al., 1997; Hillenbrand and Houde, 1996; McHenry, 1999; Sapienza et al., 1999).

The *Goldilocks Passage* was written to include words and segments that would be likely to reveal dialect variation. This passage has been used in the United Kingdom in sociolinguistic studies of language variation and attitudes (Stockwell, 2002).

Finally, each talker was recorded while engaged in two fluent conversations with the experimenter. One of the conversations was 5 min in length and included questions about the talker's hometown, extracurricular activities, and travel experiences. The other conversation varied from 7 to 12 min in length and was designed to elicit certain target words from the talker in relatively natural, conversational speech. Through a series of questions related to specific topics, 10 target monosyllabic words, each containing a different vowel, were elicited from each talker.¹

2.2. Talkers

Sixty white talkers between the ages of 18 and 25 years old were recruited from the Indiana University community for participation in the NSP corpus. All of the talkers were monolingual native speakers of American English with no history of hearing or speech disorders reported at the time of testing. Both parents of each talker were also native English speakers. Most of the talkers ($N = 54$) were undergraduates at the time of recording, but five had completed a bachelor's degree and one had completed a master's degree. The socioeconomic status of the participants, as indicated by the occupation and level of education of the participants' parents, was somewhat varied, although in most cases ($N = 50$) at least one parent had a bachelor's degree and only three participants had two parents with no undergraduate education. Thus, most of the talkers were from middle or upper-middle class backgrounds.

The 60 talkers included five males and five females from each of six dialect regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West. These six regions were selected based on Labov et al.'s (forthcoming) dialect categories (see Fig. 1). In order to ensure that we would obtain a large enough sample of talkers from the Indiana University community while geographically covering as much of the United States as possible, we combined Eastern and Western New England into a single New England category for the purposes of the corpus and we did not attempt to collect data from talkers who came from Western Pennsylvania or Florida.

In order to reduce the effects of dialect leveling, each talker had lived in Bloomington, Indiana, for less than 2 years at the time of recording and had lived in a single target dialect region for his or her entire life prior to moving to Bloomington. Both parents of each talker were also raised in that same target dialect region. The map in Fig. 4 shows the hometowns for each of the 60 NSP talkers. Male talkers are represented by dark dots and female talkers are represented by light squares. As shown in Fig. 4, the hometowns of the talkers are unevenly distributed geographically within each region. In particular, most of the Northern talkers are from northern Indiana and northern Illinois, most of the Midland talkers are from central Indiana, half of the Southern talkers are from the Louisville, Kentucky metropolitan area, and half of the Western talkers are from southern California. However, both Eastern and Western New England are represented in the New England talkers and the Mid-Atlantic is represented by talkers from New York, New Jersey, and Maryland. This geographic distribution reflects the undergraduate student population at Indiana University; based on the hometowns provided by incoming freshman at Indiana University in 2002, $\approx 64\%$ of the

¹The authors would like to thank Nancy Niedzielski for suggesting this targeted interview task as a means of eliciting specific words in relatively natural continuous speech.

students from western states are from California, $\approx 25\%$ of the students from southern states are from Kentucky, and $\approx 66\%$ of the total undergraduate population is from Indiana.

2.3. Procedures

Participants were recorded one at a time by the first author (CGC) in a sound attenuated booth (IAC Audiometric Testing Room, Model 402). Both the experimenter and the participant sat in the sound booth during testing. During the recording session, the participant was seated in front of a ViewSonic LCD flatscreen monitor (ViewPanel VG151) which mirrored the screen of a Macintosh Powerbook G3 laptop. The participant wore a Shure head-mounted microphone (SM10A) that was positioned approximately one inch from the left corner of the talker's mouth. The microphone output was fed to an Applied Research Technology microphone tube pre-amplifier. The output gain on the pre-amplifier was adjusted by the experimenter while the participant read the *Grandfather Passage* (Darley et al., 1975) as a warm-up before recording began. The output of the microphone pre-amplifier was connected to a Roland UA-30 USB audio interface which digitized the signal and transmitted it via USB ports to the laptop where each utterance was recorded in an individual AIFF 16-bit digital sound file at a sampling rate of 44.1 kHz. The experimenter held the laptop on her lap and wore headphones connected to the Roland device so that she could hear the same audio signal that was being input to the laptop for recording.

The presentation of the stimulus materials was controlled by the Macintosh laptop using home-grown software. Stimulus items were presented one at a time in 24-point green Courier font on a black background on both the laptop on the experimenter's lap and the LCD screen in front of the participant. The stimulus materials were presented in blocks, such that each participant read all of the CVC words in one experimental block, all of the high probability sentences in another block, and so on, for a total of 10 experimental blocks. Prior to the beginning of each block of trials, the participant was given written instructions on the LCD screen and verbal instructions by the experimenter. Participants were permitted to take breaks between blocks as needed. Each participant within a given dialect received the stimulus materials in a different, random order.

The stimulus items were presented one at a time in random order on the laptop and LCD screens. The duration of the recording intervals varied with each stimulus type and are shown in Table 4. If the participant misread an item or if there was any background noise while the participant read the item, the trial was repeated at the end of the experimental block.

The entire recording session lasted ≈ 1 h. The participants received \$15 in payment and a Speech Research Laboratory t-shirt for their service.

The corpus of digital speech samples is organized by talker and then by stimulus type in a hierarchical file structure. Each stimulus item is identified by a three character talker identifier, a one character stimulus type identifier, and a four character stimulus number. A definition file is available for each talker which lists the order of the presentation of the experimental blocks for that talker. In addition, for each set of materials for each talker, a log file is available which provides the order of presentation of the stimulus items for that talker, as well as links between the stimulus number, the orthographic transcription of the stimulus, and the audio file name. Finally, orthographic transcriptions are available for the two interview speech samples from each talker.

3. Research applications of the NSP corpus

Materials from the NSP corpus have recently been used in several acoustic and perceptual experiments. A preliminary acoustic analysis of the hVd productions confirmed significant differences in vowel production due to regional dialect (Clopper et al., 2005b) and a factor analysis of the results suggested that the Northern Cities Chain Shift and Southern Vowel Shift are two of the most prominent acoustic properties that characterize regional varieties of American English (Clopper and Paolillo, 2005). Fig. 5 shows the mean vowel spaces for each of the six dialects, based on five hVd utterances produced by four male and four female talkers from each of the six regions. The Northern Cities Chain Shift is clearly present in the speech of the Northern talkers, particularly with respect to the raising and fronting of /æ/ and the fronting of /ɑ/. In addition, the Southern talkers produced aspects of the Southern Vowel Shift, including the fronting of /u/ and /o/, the raising of /ɛ/ and the lowering of /e/. Fig. 5 also shows /u/ fronting in the Western and Midland dialects as well as a partial merger of /ɑ/ and /ɔ/ in the New England, Midland, and Western dialects. Thus, variation in vowel production due to regional dialect is present in the speech of the NSP talkers even in the most constrained set of stimulus materials, the hVd words. We expect that further acoustic analyses of the other materials will provide additional insights into the range of variation produced by this sample of talkers.

The high probability sentence materials have been used as stimulus materials in a series of perceptual tasks, including dialect categorization (Clopper, 2004), auditory free classification of dialects (Clopper, 2004), and paired comparison dialect similarity ratings tasks (Clopper et al., in press). The results of all three perceptual tasks confirmed that naïve listeners could perceive and represent salient acoustic–phonetic properties of the different regional varieties of American English and then make explicit judgments about the talkers based on regional dialect. The recordings included in the NSP corpus will therefore be useful for a wide range of speech science and sociolinguistic investigations of the perception of linguistic variation.

4. Conclusions

The Nationwide Speech Project is a new corpus containing recordings of 60 young adult talkers representing six different regional varieties of American English: New England, Mid-Atlantic, North, Midland, South, and West. The speech samples obtained from each talker include isolated words, sentences, passages, and samples of interview speech. Given the results that have already been obtained in acoustic and perceptual experiments using stimulus materials from the NSP corpus, we expect that the corpus will be useful for a wide range of novel perceptual and acoustic experiments designed to explore the role of variation in spoken language processing. Copies of the corpus will be made available to the research community upon request from the first author (CGC). Portions of the NSP corpus will also be distributed through the Linguistic Data Consortium.

Acknowledgments

This work was supported by NIH NIDCD T32 Training Grant DC00012 and NIH NIDCD R01 Research Grant DC00111 to Indiana University. The authors would like to acknowledge the contributions of Allyson Carter, Connie Clarke, Caitlin Dillon, Jimmy Harnsberger, Rebecca Herman, and Luis Hernandez to the development of the Nationwide Speech Project corpus, including the compilation of the stimulus materials, the selection of the equipment, and pilot testing of both equipment and participants.

References

- Ash, S. Sampling strategy for the Telsur/Atlas project. n.d. Available from: <http://www.ling.upenn.edu/phono_atlas/sampling.html>
- Baker K, Ramig L, Johnson A, Freed C. Preliminary voice and speech analysis following fetal dopamine transplants in five individuals with Parkinson disease. *J Speech Lang Hear Res.* 1997; 40:615–626. [PubMed: 9210118]
- Boberg C. The phonological status of Western New England. *Amer Speech.* 2001; 76:3–29.
- Byrd D. Relations of sex and dialect to reduction. *Speech Comm.* 1994; 15:39–54.
- Canavan, A.; Zepperlen, G. CALLFRIEND American English-Non-Southern Dialect (LDC96S46). Linguistic Data Consortium; Philadelphia: 1996a.
- Canavan, A.; Zepperlen, G. CALLFRIEND American English-Southern Dialect (LDC96S47). Linguistic Data Consortium; Philadelphia: 1996b.
- Carter AK, Clopper CG. Prosodic effects on word reduction. *Lang Speech.* 2002; 45:321–353. [PubMed: 12866908]
- Carver, CM. *American Regional Dialects: A Word Geography.* University of Michigan Press; Ann Arbor, MI: 1987.
- Clopper, CG. Unpublished doctoral dissertation. Indiana University; Bloomington: 2004. Linguistic experience and the perceptual classification of dialect variation.
- Clopper, CG.; Paolillo, JC. North American English vowels: A factor-analytic perspective. Paper presented at *Methods in Dialectology XII*; Moncton, NB. 2005.
- Clopper CG, Pisoni DB. Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Lang Var Change.* 2004a; 16:31–48.
- Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *J Phonetics.* 2004b; 32:111–140.
- Clopper, CG.; Carter, AK.; Dillon, CM.; Hernandez, LR.; Pisoni, DB.; Clarke, CM.; Harnsberger, JD.; Herman, R. Research on Spoken Language Processing Progress Report No. 25. Speech Research Laboratory, Indiana University; Bloomington, IN: 2001. The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus; p. 367-380.
- Clopper CG, Conrey BL, Pisoni DB. Effects of talker gender on dialect categorization. *J Lang Soc Psychol.* 2005a; 24:182–206.
- Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *J Acoust Soc Amer.* 2005b; 118:1661–1676. [PubMed: 16240825]
- Clopper CG, Levi SV, Pisoni DB. Perceptual similarity of regional dialects of American English. *J Acoust Soc Amer.* in press.
- Darley, FL.; Aronson, AE.; Brown, JR. *Motor Speech Disorders.* Saunders; Philadelphia: 1975.
- Dictionary of American Regional English, 1985.* Vol. 1 (A–C), Cassidy, F.G. (Ed.), Vol. 2 (D–H) and Vol. 3 (I–O), Cassidy, F.G., Hall, J.H. (Eds.), Vol. 4 (P–Sk), Hall, J.H. (Ed.). Belknap Press, Cambridge, MA.
- DuBois, JW.; Chafe, WL.; Meyer, C.; Thompson, SA. Santa Barbara Corpus of Spoken American English Part-I (LDC2000S85). Linguistic Data Consortium; Philadelphia: 2000.
- Fairbanks, G. *Voice and Articulation Drillbook.* Harper; New York: 1940.
- Feagin, C. Entering the community: Fieldwork. In: Chambers, JK.; Trudgill, P.; Schilling-Estes, N., editors. *The Handbook of Language Variation and Change.* Blackwell; Malden, MA: 2002. p. 20-39.
- Fisher, WM.; Doddington, GR.; Goudie-Marshall, KM. The DARPA speech recognition research database: Specification and status. *Proceedings of the DARPA Speech Recognition Workshop;* 1986. p. 93-99.
- Gelfer M, Schofield K. Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *J Voice.* 2000; 14:22–33. [PubMed: 10764114]
- Giles, H.; Powesland, P. Accommodation theory. In: Coupland, N.; Jaworski, A., editors. *Sociolinguistics: A Reader and Coursebook.* Palgrave; New York: 1997. p. 232-239.

- Hall, JH.; von Schneidmeyer, L. Dictionary of American Regional English. 2004. Available from: <<http://polyglot.lss.wisc.edu/dare/dare.html>>
- Hillenbrand J, Houde R. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech Hear Res.* 1996; 39:311–321. [PubMed: 8729919]
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Amer.* 1977; 61:1337–1351. [PubMed: 881487]
- Krapp, GP. *The English Language in America.* Frederick Ungar; New York: 1925.
- Labov, W. *Sociolinguistic Patterns.* University of Pennsylvania Press; Philadelphia: 1972a.
- Labov W. Some principles of linguistic methodology. *Lang Soc.* 1972b; 1:97–120.
- Labov, W. *Principles of Linguistic Change: Internal Factors.* Blackwell; Malden, MA: 1994.
- Labov, W. The three dialects of English. In: Linn, MD., editor. *Handbook of Dialects and Language Variation.* Academic Press; San Diego: 1998. p. 39-81.
- Labov, W.; Ash, S.; Boberg, C. *Atlas of North American English.* Mouton de Gruyter; New York: forthcoming
- McDavid, RI, Jr. The dialects of American English. In: Francis, WN., editor. *The Structure of American English.* Ronald Press; New York: 1958. p. 480-543.
- McHenry M. Aerodynamic, acoustic and perceptual measures of nasality following traumatic brain injury. *Brain Injury.* 1999; 13:281–290. [PubMed: 10230529]
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on Speech Perception Progress Report No. 10. Speech Research Laboratory, Indiana University; Bloomington, IN: 1984. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words; p. 357-376.
- Plichta, B.; Mendoza-Denton, N. Best practices in the digital acquisition and processing of acoustic speech data. Paper presented at *New Ways of Analyzing Variation*; Raleigh, NC. 2001.
- Rischel J. Formal linguistics and real speech. *Speech Comm.* 1992; 11:379–392.
- Rojas, DM. PIQUANT-LA: Probabilistic Identification and Quantification of Linguistic Affinity. University of Edinburgh M.Sc.thesis; 2002.
- Sapienza C, Walton S, Murry T. Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *J Speech Lang Hear Res.* 1999; 42:127–140. [PubMed: 10025549]
- Stockwell, P. *Sociolinguistics: A Resource Book for Students.* Routledge; London: 2002.
- Strassel, S.; Conn, J.; Wagner, SE.; Cieri, C.; Labov, W.; Maeda, K. SLX corpus of Classic Sociolinguistic Interviews (LDC2003T15). Linguistic Data Consortium; Philadelphia: 2003.
- Thomas, ER. *An Acoustic Analysis of Vowel Variation in New World English.* Duke University Press; Durham, NC: 2001.
- Trudgill, P. Accommodation between dialects. In: Linn, MD., editor. *Handbook of Dialect and Language Variation.* Academic Press; San Diego: 1998. p. 307-342.
- Wolfram, W.; Fasold, RW. Field methods in the study of social dialects. In: Coupland, N.; Jaworski, A., editors. *Sociolinguistics: A Reader and Coursebook.* Palgrave; New York: 1997. p. 89-115.
- Zue V, Seneff S, Glass J. Speech database development at MIT: TIMIT and beyond. *Speech Comm.* 1990; 9:351–356.



Fig. 1.
The major dialects of American English, based on Labov et al. (forthcoming).

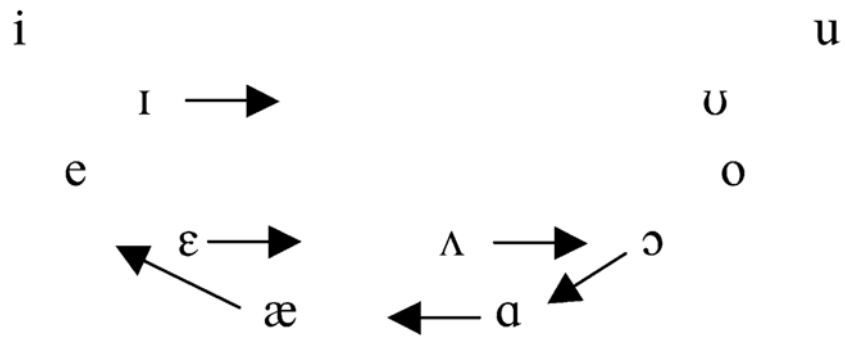


Fig. 2.
The Northern Cities Chain Shift.

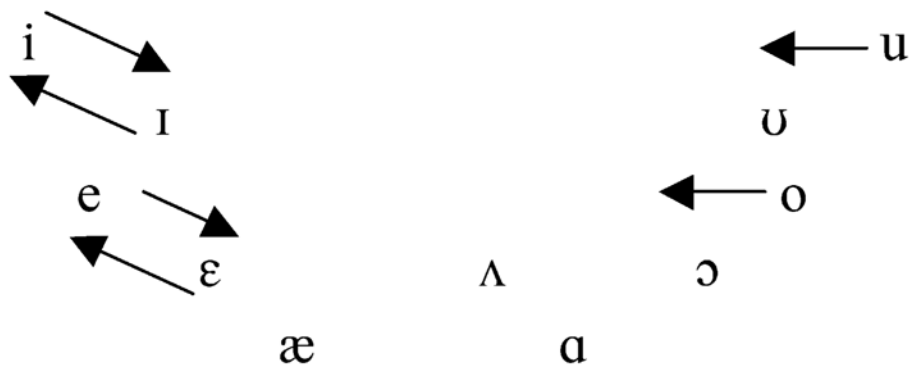


Fig. 3.
The Southern Vowel Shift.



Fig. 4. Map of the hometowns of the 60 talkers included in the Nationwide Speech Project corpus. Dark dots represent male talkers and light squares represent female talkers.

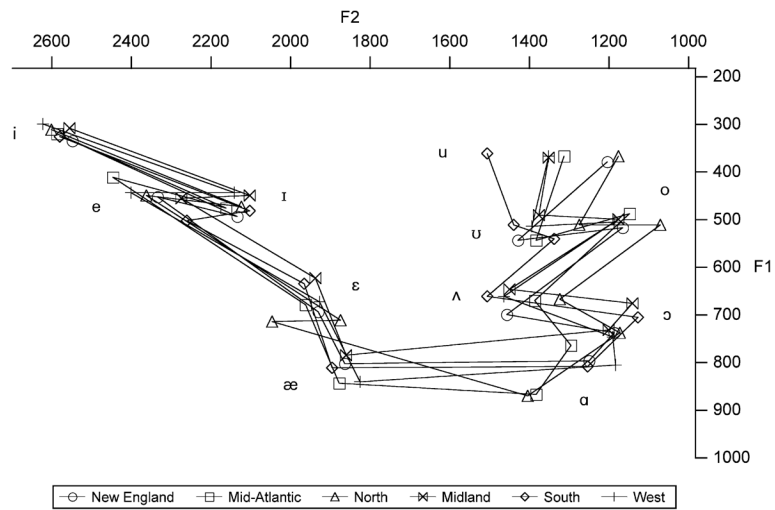


Fig. 5.

Mean vowel spaces for each of the six dialect regions included in the NSP corpus. The vowels proceed from left to right: i, ɪ, e, ε, æ, α, ə, ʌ, o, ɔ, u. Each data point represents the average first and second formant frequencies of the vowel obtained from five hVd productions from each of four male and four female talkers in the NSP corpus. Replotted from (Clopper et al., 2005b).

Table 1

Some factors to consider in designing a corpus of spoken language

| Factor | Examples |
|--------------------------|------------------------------------------------------------------------------------------------------------------------|
| Talker demographics | Age, gender, socioeconomic status, race, ethnicity, level of education, residential history, linguistic experience |
| Interviewer demographics | Insider vs. outsider |
| Recording conditions | Fieldwork recordings on tape (analog or digital), telephone recordings, digital recordings in a sound-attenuated booth |
| Speech materials | Spontaneous speech, interview speech, read speech ("lab speech") |

Table 2

Summary of existing speech corpora with regional dialect variation

| Corpus | Description |
|---------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dictionary of American Regional English (Hall and von Schneidmeyer, 2004) | <ul style="list-style-type: none"> • Fieldwork interviews in 1000 communities in the United States • 1800-item interview questionnaire and a read passage • Used to create the five volume Dictionary of American Regional English |
| SLX Corpus of Classic Sociolinguistic Interviews (Strassel et al., 2003) | <ul style="list-style-type: none"> • 8 sociolinguistic interviews with a total of 9 talkers • Narratives, interview responses, word lists • Used as a pedagogical tool for training sociolinguistic fieldworkers |
| Santa Barbara Corpus of Spoken American English (DuBois et al., 2000) | <ul style="list-style-type: none"> • Hundreds of recordings of “natural speech” • Settings include conversations, political speeches, classroom lectures, and bedtime stories • Used for studies of prosodic variation and discourse analysis |
| Call Friend Corpora (Canavan and Zeperlen, 1996a,b) | <ul style="list-style-type: none"> • 60 telephone conversations between two American non-Southerners • 60 telephone conversations between two American Southerners • Used for spoken word recognition and automatic language identification |
| TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher et al., 1986) | <ul style="list-style-type: none"> • 630 talkers ranging in age, ethnicity, gender, education, and regional background • Read sentence materials • Used for spoken word recognition, acoustic analysis of regional variation, perceptual dialect categorization, and automatic dialect identification |

Table 3

The speech materials collected from each talker in the NSP corpus

| | Materials set | Number of tokens | Examples |
|------------------|----------------------------|-------------------------|----------------------------------------------------------------------|
| Words | hVd words | 10 | heed, hid, head |
| | CVC words | 76 | mice, dome, bait |
| | Multi-syllabic words | 112 | alfalfa, nectarine |
| Sentences | High probability sentences | 102 | Ruth had a necklace of glass beads The swimmer dove into the pool |
| | Low probability sentences | 52 | Tom has been discussing the beads She might consider the pool |
| | Anomalous sentences | 52 | Bill knew a can of maple beads The jar swept up the pool |
| Passages | Rainbow passage | 1 | When sunlight strikes the raindrops in the air ... |
| | Goldilocks passage | 1 | Once upon a time, there were three bears ... |
| Interview speech | Interview speech | (5 min) | hometown, travel experiences |
| | Targeted interview speech | 10 | sleep, shoes, math |

Table 4

Experiment specifications for the NSP corpus

| Materials set | Recording time (s) | Inter-trial interval (s) |
|----------------------------|---------------------------|---------------------------------|
| hVd words | 2 | 0.5 |
| CVC words | 2.25 | 0.5 |
| Multi-syllabic words | 3.5 | 0.5 |
| High probability sentences | 5 | 0.5 |
| Low probability sentences | 5 | 0.5 |
| Anomalous sentences | 6 | 0.5 |
| Rainbow passage | Untimed | None |
| Goldilocks passage | Untimed | None |
| Interview speech | 300 (5 min) | None |
| Targeted interview speech | Untimed | None |