RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 25 (2001-2002)
Indiana University

The Indiana Speech Project: An Overview of the Development of a Multi-Talker Multi-Dialect Speech Corpus[1]

Cynthia G. Clopper, Allyson K. Carter, Caitlin M. Dillon, Luis R. Hernandez, David B. Pisoni, Connie M. Clarke,[2] James D. Harnsberger,[3] and Rebecca Herman[4]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Research Grant DC00111 to Indiana University.
[2] University of Arizona, Tucson, AZ.
[3] University of Florida, Gainesville, FL.
[4] DePaul University, Chicago, IL.

The Indiana Speech Project: An Overview of the Development of a Multi-Talker Multi-Dialect Speech Corpus

Abstract. The goal of the Indiana Speech Project (ISP) was to collect a corpus of spoken language samples from a number of different talkers that represent several different regional varieties of American English. Audio recordings were made of five college-aged women from each of six geographic regions of Indiana while they read isolated words, sentences, and a passage and while engaged in a conversation with an experimenter. The residential histories of the women and those of their parents were strictly controlled to ensure that the talkers were good representatives of each dialect region. The Indiana speech corpus will be used for acoustic-phonetic measurements of speech and perceptual studies on regional language variation by different groups of listeners.

Introduction and Theoretical Motivation

In recent years, researchers in the field of speech perception and spoken language processing have developed a number of speech corpora for conducting acoustic and perceptual experiments on human language. These corpora typically included a number of speakers reading a set of words or sentences [e.g., "Easy-Hard" Word Multi-Talker Speech Database (Torretta, 1995); TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990); Talker Variability Sentence Database (Karl & Pisoni, 1994)]. Although the talkers often included both males and females, other important indexical variables such as socioeconomic status, age, ethnicity, and regional dialect were rarely, if ever, considered in selecting the talkers. One exception to this rule is the TIMIT Acoustic-Phonetic Continuous Speech Corpus. This corpus contains spoken sentence materials from 630 talkers, representing eight different regional dialects of American English (Zue et al., 1990). While the TIMIT database was originally collected for speech recognition research, it has been used in various acoustic-phonetic studies on the role of gender, age, and dialect in linguistic variation (e.g., Byrd, 1992; Byrd, 1994; Keating, Blankenship, Byrd, Flemming, & Todaka, 1992; Keating, Byrd, Flemming, & Todaka, 1994).

While this work was going on in speech perception and speech recognition, sociolinguists have been collecting speech samples from talkers of a variety of ages, socioeconomic statuses, ethnicities, and regional dialects. The emphasis in this research has been on capturing the variability in spoken language as well as collecting extensive demographic information on each talker [e.g., Santa Barbara Corpus of Spoken American English (DuBois, Chafe, Meyer, & Thompson, 2000); CallFriend Telephone Speech Corpus for American English (Linguistic Data Consortium, 1996); TELSUR (Labov, Ash, & Boberg, in press)]. In contrast to the speech stimuli used in speech perception and speech recognition research, these speech samples are typically taken from "natural" language situations such as interviews and telephone calls and less emphasis is placed on obtaining identical utterances from the same set of talkers. While these corpora are useful for many kinds of sociolinguistic research, they are not adequate for perceptual research in which consistent linguistic content across talkers is highly desirable.

The initial goal of the Indiana Speech Project (ISP) was to collect a large amount of speech from a number of phonologically distinct dialect regions in the state of Indiana for use in perceptual studies and acoustic analyses. We wanted to combine the best aspects of the speech perception corpora with the unique focus of the sociolinguistic corpora. In particular, our goal was to collect a large corpus of utterances that were consistent across all talkers, allowing for better control of the stimulus materials for a wide range of perceptual and acoustic studies. In addition, we also wanted the talkers in our corpus to be homogeneous on all variables except regional dialect, which we manipulated following the strict guidelines described in detail below.

The TIMIT corpus was originally collected for speech recognition research, which means that it lacks two important elements for speech perception research. First, while each talker in the TIMIT corpus was recorded reading ten sentences, only two test sentences were spoken by all 630 talkers. As a result, researchers interested in comparing talkers without introducing variation in linguistic content are limited to just two sentences. Second, each talker in the TIMIT corpus was assigned to one of eight regional dialect labels, but the precise details of how those assignments were made is not explained in the accompanying documentation. In general, the dialect labels were assigned based on where the talker spent the majority of his or her childhood (William Fisher, personal communication), but there was little, if any, specific information collected about the regional dialect of the talker's parents. In addition, the labels used to define the dialects of the TIMIT talkers and the map used to define the regions (see Fisher, Doddington, & Goudie-Marshall, 1986) does not correspond with any current sociolinguistic theory of American English regional dialect variation.

Our speech corpus was designed to be an improvement over the TIMIT corpus because it was designed specifically for perceptual and acoustic research on dialect variation in American English. First, the ISP corpus was designed such that every talker was recorded reading the same materials with the exception of a short spontaneous speech sample. Second, the talkers in the ISP corpus were carefully screened on a number of variables related to their own residential history, as well as that of their parents, in order to ensure that the dialect labels they were assigned were appropriate. Thus, the two major features of this new corpus are the controlled nature of the stimulus set and the procedures used to carefully screen, select, and assign the talkers to dialect categories.

The ISP corpus will allow us to continue and extend the research on the perception of dialect variation that we have been conducting using the TIMIT corpus (e.g., Clopper, 2000; Clopper & Pisoni, 2002). In particular, this new corpus will allow us to further investigate dialect identification and categorization, dialect discrimination by both native and non-native speakers, the role of dialect variation in lexical decision and word recognition studies, dialect intelligibility, and the role of dialect variation in perceptual learning of novel voices and novel dialects. In addition, the controlled nature of the stimulus items we have recorded will allow us to conduct more detailed acoustic analyses of the vowel systems of different regional varieties, including the study of diphthongs, and to consider some of the acoustic correlates of stress as a function of dialect.

Talkers

Thirty adult women were recruited from the Bloomington campuses of Indiana University and Ivy Tech State College to serve as talkers for the ISP corpus. All of the women were monolingual, native speakers of American English with no history of hearing or speech disorders reported at the time of recording. The women ranged in age from 18 to 22 years old, with a mean age of approximately 20. Both the mother and the father of each participant were also native speakers of American English. Twenty-nine of the 30 talkers were Caucasian (non-Hispanic) and the remaining talker was Hispanic in ethnicity but was not fluent in Spanish. All of the talkers also had limited experience with foreign languages (typically restricted to classroom instruction). They also had little, if any, formal knowledge of linguistics or linguistic research methods. All participants were paid $25 ($10/hour) and were given a free Speech Research Laboratory t-shirt for their time.

Female talkers were selected for use in this corpus for several reasons. First, women are much more cooperative as participants in studies of this kind that require several sessions of an hour to an hour and a half in length. Second, sociolinguists have found that women tend to be ahead of men in language changes in progress, regardless of whether the changes are above or below the level of conscious social awareness (Labov, 1990; Milroy & Milroy, 1993). Speech stimuli recorded from female talkers might therefore be expected to reveal current changes in progress, such as the Northern Cities shift in Chicago or the Southern vowel shift in Louisville.

The talkers were specifically recruited to represent six different dialect regions of Indiana: Fort Wayne, Indianapolis, Bloomington, Evansville, and the Indiana counties near Chicago, IL and Louisville, KY. The map in Figure 1 shows the state of Indiana and its 92 counties. The regions were defined as the county containing the city of interest and all contiguous counties. In the case of Louisville, the region was defined as those counties in Indiana that are contiguous with the county in Kentucky in which Louisville is located. In the case of Chicago, the region was defined as the three counties in Indiana that, like Chicago, are on Central Time. No participants were actually from either Chicago or Louisville. In Figure 1, the county in which the city of interest is located is shown in dark gray and the regions are shown in light gray. All participants were required to have lived exclusively in the region they were representing prior to attending school in Bloomington. In order to participate, both of the talker's parents had to have grown up in a county contiguous with or within the region of interest. For example, a participant from Evansville (located in Vanderburgh County) had to live exclusively in Posey, Vanderburgh, Warrick, and/or Gibson counties and both of her parents had to have grown up in any of those counties and/or Spencer, DuBois, Pike, or Knox counties.

Regional varieties of the English spoken in Indiana were selected for the ISP for several reasons. First, given our location in Bloomington, Indiana at Indiana University, we had access to a large number of undergraduate students from all over the state of Indiana. Second, Indiana has historically been more diverse linguistically and culturally than its neighbors on either side, Illinois and Ohio. In particular, an examination of the settlement patterns in the early 19th century reveals that roughly the southern half of Indiana was first settled by people migrating from Kentucky and points farther south who followed the Ohio River and its tributaries into southern Indiana (Bergquist, 1981). The linguistic result of this settlement pattern is the extension of Southern English dialect features in both phonology and the lexicon into much of southern Indiana (Carmony, 1965; Gibbens, 1962). Because such Southern English features are not found in southern Illinois or southern Ohio, the northward shift of the major north-south dialect boundary in Indiana is called the "Hoosier Apex" (Carver, 1987).

We therefore expected to find sizable differences in pronunciation patterns between northern and southern Indiana speakers as a result of the major north-south boundary that runs through Indiana. This major dialect boundary should separate Chicago, Fort Wayne, and Indianapolis from Bloomington, Evansville, and Louisville. In addition, Chicago has long been identified as one of the cities in which the Northern Cities vowel shift is taking place (Labov, Yeager, & Steiner, 1972), so we expected the Chicago suburbs in Indiana to reflect this change as well. At the other extreme, we predicted that the most southern features would be found in the speech of the women from the Louisville suburbs, given their proximity to Kentucky. We expected the remaining cities to fall on a continuum of variation between these two endpoints, reflecting a gradual shift from northern to southern varieties of American English (Davis & Houck, 1992).

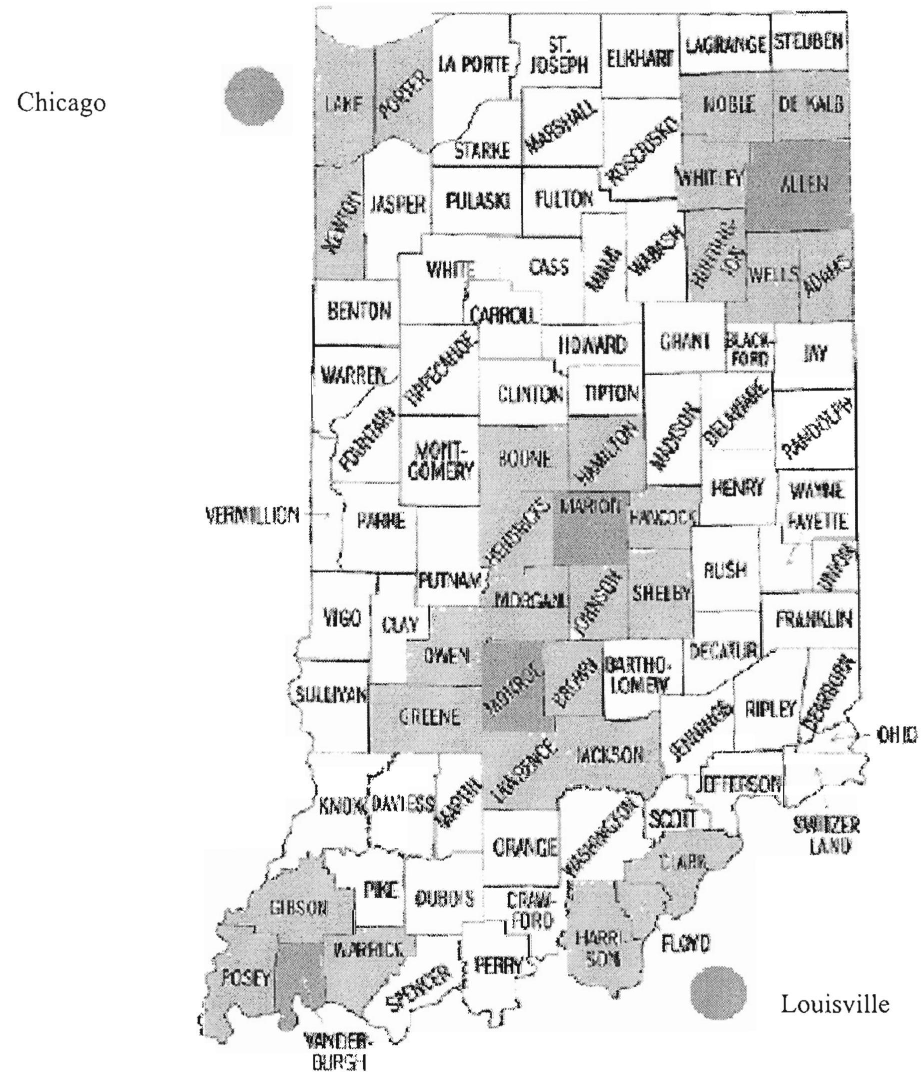

Figure 1. Map of the state of Indiana, including county lines. The dark counties represent those containing a city of interest: Allen (Fort Wayne), Marion (Indianapolis), Monroe (Bloomington), and Vanderburgh (Evansville). The dark circles outside the state line represent the other two cities of interest: Chicago, IL (to the northwest) and Louisville, KY (to the southeast). The light gray counties are those that belong to the region surrounding one of the six key cities.

Materials

Each talker was recorded while she read eight different sets of test materials. Recordings were also made of a spontaneous speech sample obtained from a short conversation with the experimenter. These materials are shown in Table 1 and are described in more detail in the sections below.

Materials Set	Number of Tokens	Examples
CVC Words	991	mice, dome, bait
Multisyllabic Words	240	alfalfa, nectarine
Disyllabic Nonwords	56	deploze, soggith
High Probability SPIN Sentences	200	All the flowers were in bloom. For your birthday I baked a cake.
Low Probability SPIN Sentences	100	Ruth will consider the herd. David does not discuss the hug.
Anomalous Sentences	100	Bill knew a can of maple beads. We're super so let's fish in the map.
Vowel Space	10 (10 repetitions)	heed, hid, head
Rainbow Passage	1	
Spontaneous Speech	1 (5 minutes)	

Table 1. Spoken language materials recorded by each talker for the ISP.

CVC Words. In order to study the acoustic properties of segmental variation in as many potentially relevant linguistic contexts as possible, it was necessary to collect a large number of consonant-vowel-consonant (CVC) utterances. The CVC list was composed of 915 different CVC words selected from an online dictionary containing approximately 20,000 entries based on Webster's Pocket Dictionary (Luce & Pisoni, 1998). The list is composed of all of the CVC words in the dictionary that received an average familiarity rating of 6.0 or greater (on a 7-point scale) by undergraduates (Nusbaum, Pisoni, & Davis, 1984). From the full set of these CVC words, 76 were selected to be produced twice, for a total of 991 tokens in the CVC list. The "repeated CVC" subset was selected such that every monophthong in English occurred at least five times in the list and every diphthong occurred at least four times in the list. Additionally, all diphthongs that participate in a shift or merger in one of the major regional dialects of American English occurred at least five times (see Labov et al., in press and Thomas, 2001 for discussions of vocalic variants and mergers in American English). All of the vowels with the exception of /ɔɪ/ occurred at least five times in this repeated CVC list. The final consonants in these words were selected such that contexts of dialect interest for each vowel were represented at least once. Contexts of interest for a given vowel were based on expected dialectal variation due to documented shifts and mergers (see Labov et al., in press and Thomas, 2001 for documentation of vocalic variation in American English). Following consonants were also varied systematically to include voiceless stops, voiced stops, nasals, liquids, and sometimes fricatives. The CVC words that were produced twice in the CVC list are shown in Appendix 1.

Multisyllabic Words. In order to study the effects of segmental interactions due to consonant clusters and stress, we also included a short list of multisyllabic words in the corpus. The multisyllabic word list consisted of a subset of the test items developed by Carter and Clopper (2000) for studying word reduction behavior in adult populations. The list was balanced for number of syllables, primary stress location, and morphological complexity. Table 2 shows the distribution of the words in this list with respect to these three variables. The list contained eight sets of 30 words of two-, three-, and four-syllables with primary stress equally distributed on the first, second, and third syllable. Half of the words in each set were monomorphemic and half were polymorphemic. All of the words were randomly selected from an online dictionary containing approximately 20,000 entries based on Webster's Pocket Dictionary. In addition, all of the words received a familiarity rating of at least 6.0 (on a 7-point scale) from undergraduate students, had a lexical frequency rating of one standard deviation from the log mean frequency, and had a neighborhood density of 2 or lower (Nusbaum et al., 1984). These criteria mean that the multisyllabic words were highly familiar, commonly occuring English words with few phonologically similar words that they might be confused with.

Number of Syllables	Primary Stress Location	Morphological Complexity	Number of Words
2	1st	monomorphemic	15
		polymorphemic	15
	2nd	monomorphemic	15
		polymorphemic	15
3	1st	monomorphemic	15
		polymorphemic	15
	2nd	monomorphemic	15
		polymorphemic	15
	3rd	monomorphemic	15
		polymorphemic	15
4	1st	monomorphemic	15
		polymorphemic	15
	2nd	monomorphemic	15
		polymorphemic	15
	3rd	monomorphemic	15
		polymorphemic	15

Table 2. Distribution of words in the multisyllabic word list with respect to number of syllables, primary stress location, and morphological complexity.

Disyllabic Nonwords. A set of disyllabic nonwords was also included in the corpus in order to compare American English vowels in the stressed syllable position of words with two different stress patterns, initial stress (Strong-weak or Sw) and final stress (weak-Strong or wS). The nonwords consisted of 56 disyllables representing fourteen American English vowels, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ʌ/, /a/, /u/, /ʊ/, /oʊ/, /ɔ/, /ai/, /aʊ/, and /oɪ/. All of these vowels except /oʊ/, /ai/, /aʊ/, and /oɪ/ were placed in those conditioning contexts in which particular dialect variants or mergers occur (see Labov et al., in press and Thomas, 2001 for discussions of vocalic variants and mergers in American English).

The 56 nonwords were designed to elicit the two different stress patterns (initial or final stress) by manipulating their syllabic structure. Twenty-eight of the nonwords (two tokens of each vowel) were designed to elicit initial stress ("Sw" words) by taking the form of a CVCVC sequence in which the medial C was always ambisyllabic. Such a "balanced" sequence of consonants and vowels is predicted to elicit initial stress because of stress biases in the American English lexicon. For example, in a study by Cutler and Carter (1987) of a corpus of 190,000 English words, 90% of content words were found to begin with "strong" syllables (e.g., clóset, óctopus, and sálamànder), which correspond to initial stress in our disyllabic nonwords.

The other half of the nonword stimulus set consisted of 28 disyllabic nonwords (two tokens of each vowel) designed to elicit final stress ("wS" words). Given the bias in English for disyllabic words to bear initial stress, it was critical to make use of certain phonological and syntactic patterns in English in order to elicit a consistent final stress pronunciation. First, CVCCVC sequences were used for all nonwords with predicted final stress. The medial consonant clusters in these sequences were always permissible word onset clusters in English, to encourage listeners to parse the nonwords as a CV.CCVC. Given the role that syllable weight normally plays in stress assignment in English and other languages, this parsing should elicit second syllable stress for these words (Hammond, 1999; Hayes, 1995).

Finally, for all nonwords, two additional criteria were adopted to ensure that the stimuli were sufficiently different from any real English words and that they would be produced with the intended vowel qualities and stress assignments. First, neither syllable in any of the disyllabic nonwords constituted a real English word orthographically (including proper names, but not slang). Second, all syllables involved phonotactically permissible segment sequences in English.

To encourage the appropriate stress assignment in production, the Sw and wS nonwords were placed in sentential contexts that were predicted to favor either initial or final stress, respectively. It has been shown experimentally that grammatical position affects stress assignment in nonwords (Kelly, 1988; Kelly & Bock, 1988). In one experiment, participants were presented with disyllabic nonsense words in various sentential contexts. The results showed that the talkers were more likely to pronounce nonsense words in noun contexts with a Sw pattern (Kelly and Bock, 1988). In another experiment, participants were presented with tape-recorded disyllabic nonwords and asked to construct sentences with these items (Kelly, 1988). Participants were significantly more prone to place words with final stress (wS) in verbal positions. Given these results, the nonwords designed to elicit initial stress were placed in a noun context, namely the subject position in the carrier sentence, "The ____ chased the ball." The nonwords designed to elicit final stress were placed in a verb context in the carrier sentence, "He will _____ the cookie dough."

Development of the nonword stimuli required two pilot experiments, using a total of 23 participants. In the first pilot, 13 participants were asked to produce the first iteration of the 56 nonwords in the carrier sentences discussed above. The 56 trials were randomized across initial stress and final stress. Analysis of stress location accuracy revealed that 40 of the 56 forms were produced with a stress accuracy rate of 86% or more, while 16 were produced with a stress accuracy rate of 78% or less. We then conducted a second pilot experiment in which we changed the orthography of the 16 words that were below the accuracy criterion of 86% to avoid problematic segmental effects that seemed to be common to all 16 below-criterion forms. We also changed the procedure such that trials were presented in two blocks. Subjects were first presented with all nonwords with predicted final stress, followed by all nonwords with predicted initial stress. This ordering was chosen to discourage talkers from carrying over the preferred initial stress pattern from noun context sentences to verb context sentences (with predicted final stress nonwords). Within blocks, individual trials were randomized. Ten participants were presented with all 40 of the "old" nonwords that had an accuracy rate over 86%, as well as the new, altered set of 16 nonwords. Results showed that each of the "new" nonwords reached an accuracy rate above the 86% criterion. The disyllabic nonword list consisted of these 56 nonwords from the two pilots and is shown in Appendix 2. Examples of the nonwords are shown in Table 1.

High and Low Probability SPIN Sentences. In addition to isolated words, we also included three sets of sentences for use in acoustic and perceptual studies of variation. Two of the three sentence lists were taken from the Speech Perception in Noise (SPIN) test (Kalikow, Stevens, & Elliot, 1977). These sentences are 5 to 8 words in length and are phonetically balanced based on phoneme frequency in English. The final word in each of the sentences is the target word. There are eight SPIN lists composed of 50 sentences each. Twenty-five of the sentences in each list are termed high probability (HP) because the target word is predictable from the semantic context of the sentence. The other 25 sentences in each list are termed low probability (LP) because the target word in each of these sentences is not predictable from the meaning of the sentence. All 200 of the high probability SPIN sentences were recorded in the ISP corpus. Half of the low probability sentences were chosen for the ISP. In particular, all 25 low probability sentences from lists 1, 2, 7, and 8 of the original SPIN test were selected, for a total of 100 low probability SPIN sentences. Examples of the SPIN sentences are shown in Table 1.

Anomalous Sentences. One hundred semantically anomalous sentences were created based on the SPIN sentences (Kalikow, Stevens, & Elliot, 1977). Target words for high and low probability sentences in the SPIN test are matched across lists such that high probability target words in list 1 match low probability target words in 2 and vice versa. The same relationship holds for lists 7 and 8. Therefore, in order to create anomalous sentences with the same target words as the low probability sentences that had already been selected for the ISP, the sentence structure and target words were taken from the 25 high probability sentences in the original SPIN lists 1, 2, 7, and 8. All content words (nouns, verbs, adjectives, adverbs, and some prepositions) of the original sentences were replaced with randomly selected content words that were the same part of speech from high probability sentences in lists 3, 4, 5, and 6. The resulting 100 sentences were semantically anomalous, but syntactically correct.

In order to confirm that the anomalous sentences could be read by untrained talkers without difficulty or disfluency, the anomalous sentences were presented visually to 13 participants who were recorded reading them aloud. Seven sentences were read by more than three of the talkers with disfluencies and had to be revised. In order to confirm that the anomalous sentences were all roughly equivalent in terms of their semantic anomaly, the revised list was then presented visually to 10 participants who rated each sentence on a 7-point sensibleness/strangeness scale. The five sentences which fell outside of one standard deviation from the mean rating were revised. This second revised list was then presented visually to 11 new participants who again rated them on their sensibleness/strangeness on a 7-point scale. All of the revised sentences fell within a single standard deviation of the mean rating. This third list was presented visually to ten participants who read them aloud, to ensure that revisions had not caused any of the sentences to be more difficult to read aloud fluently. One of these revised sentences was read with disfluencies by more than three talkers and was revised once again. The final anomalous sentence list used in the ISP corpus was thus the result of several successive stages of revision from the original list due to both perception and production pilot studies. After this process was completed, the list consisted of 100 sentences with the same target words as the 100 low probability SPIN sentences. The anomalous sentences were also all roughly equal with regard to their semantic anomaly and could be produced by naïve talkers fluently. Examples of the anomalous sentences are shown in Table 1.

Vowel Space. The vowel space portion of the corpus consisted of a set of familiar English words that can serve as benchmarks for mapping out the vowel system of a given individual talker or dialect group. This data will allow us to directly compare the vowel spaces of talkers from the different regions of Indiana, while keeping linguistic context constant across all of the utterances. Vowel spaces are typically mapped in two-dimensions, corresponding to first and second vowel formant frequencies (Gerstman, 1968; Peterson & Barney, 1952). The stimulus materials for this portion of the corpus consisted of 10 monosyllabic words, *heed*, *hid*, *aid*, *head*, *had*, *hut*, *odd*, *who'd*, *hood*, and *owed*, representing ten American English vowels, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ʌ/, /a/, /u/, /ʊ/, and /oʊ/, respectively. The use of these particular test words was motivated by the relatively constant consonantal context in which the vowels of interest appear (hVd or _Vd, where *V* = vowel). Each of the ten items was repeated ten times during the vowel space block of the experiment.

Passage of Connected Speech. A short continuous speech sample was obtained from each talker as a controlled, read passage which would allow for detailed analysis of individual variability, as well as dialectal differences at a higher level of language use than isolated words or sentences. Several passages were considered, including the Arthur the Rat passage (used to elicit dialectal differences for the Dictionary of American Regional English; Cassidy, 1985) and the Grandfather Passage (used by researchers of speech motor disorders; cf. Darley, Aronson, & Brown, 1975). However, the passage selected for the current project was the initial paragraph of the Rainbow Passage shown in Appendix 3 (Fairbanks, 1940). The Rainbow Passage has been used in numerous acoustic and perceptual studies since its first publication, including investigations into speaker differences (e.g., Gelfer & Schofield, 2000), individual variability (e.g., Sapienza, Walton, & Murry, 1999), and clinical populations (e.g., Hillenbrand & Houde, 1996; McHenry, 1999; Baker, Ramig, Johnson, & Freed, 1997).

Spontaneous Speech. In addition to recording subjects reading words, sentences, and the passage, we also collected a five-minute sample of spontaneous speech from each participant while she was engaged in a conversation with the experimenter in order to provide another somewhat more "natural" comparison to the read speech collected with the other materials. The spontaneous speech sample was based on the participants' responses to questions concerning their demographic background, such as information about their hometown, family, hobbies, and interests.

Methods

Participants were recorded individually in two separate test sessions that took place on different days. The first session lasted approximately an hour and a half, including breaks. The second session lasted approximately an hour, including breaks. Each set of materials was presented as a single experimental block. In all seven of the tasks in which there was more than one trial (i.e., all but the passage and the spontaneous speech tasks), the individual items were presented in a different random order for each participant. Prior to beginning a new block of trials, the participants were given instructions and were encouraged to ask questions if they did not understand what they were expected to do. On the first day of recording, talkers first read the CVC words, followed by the vowel space, multisyllabic words, and spontaneous speech tasks. The latter three tasks were performed in random order by each participant. On the second day, the talkers read the remaining materials (all three sentence lists, the disyllabic nonwords, and the passage). The order of the tasks on the second day was varied randomly across talkers. Breaks were given within and between each experimental block, as needed.

During all of the recording blocks, the participants were seated in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402) in front of a ViewSonic LCD monitor (ViewPanel VG151) mirroring the screen of a Macintosh Powerbook G3 which the experimenter, who was also seated in the sound-attenuated chamber, held on his or her lap. The participants read the materials off of the monitor screen as they were presented, speaking into a head-mounted Shure microphone (SM10A) positioned at the left corner of the mouth less than an inch from the face. The microphone signal was amplified by a tube microphone preamplifier (Applied Research Technology) which was connected to a Roland UA-30 USB Audio Interface which digitized the amplified microphone signal. The output of the Roland UA-30 was sent to the Powerbook and to Audio-Technica headphones (ATH-M2X) worn by the experimenter so that he or she was able to hear the signal as it was being sent to the Powerbook.

For the real words and sentence materials, a single word or sentence was displayed on the screen on each trial and the participant was asked to read the item aloud. The amount of time that the participant had to respond to the test item on the screen varied for the different materials. These recording times are shown in Table 3. All recordings were digitized at a sampling rate of 44.1kHz and each item was recorded into an individual .aiff sound file on the Powerbook. If the participant made an error in reading the item or if the experimenter heard any extraneous noise through the headphones, the experimenter pressed the "r" key on the Powerbook keyboard and the trial was repeated again at the end of the block.

Materials Set	Recording Time
CVC Words	2250 ms
Multisyllabic Words	3500 ms
Disyllabic Nonwords	5000 ms
High Probability SPIN Sentences	5000 ms
Low Probability SPIN Sentences	5000 ms
Anomalous Sentences	6000 ms
Vowel Space	2000 ms
Rainbow Passage	1 minute
Spontaneous Speech	5 minutes

Table 3. Recording times for the different material sets.

A slightly different procedure was used for the disyllabic nonwords, the connected speech passage, and the spontaneous speech task. For the nonwords, all of the wS words were presented in random order before the Sw words, which were also presented in random order. On each trial, participants saw the carrier sentence with the nonword on the screen as well as the nonword in isolation and were instructed to read aloud the entire sentence and then to repeat the nonword by itself. Participants were given an example in the instructions and were asked to read the example aloud for the experimenter before beginning each nonword block. For the passage of connected speech, the participants were shown the passage along with the instructions and were asked to read it silently to themselves to familiarize themselves with it. They were then given 1 minute to read the passage aloud. The experimenter pressed the "s" key to stop the recording and continue on to the next experimental block when the participant had finished reading. For the spontaneous speech sample, the participant was asked to engage in a conversation with the experimenter. The experimenter asked the participant questions about her hometown and the participant was encouraged to speak freely for five minutes.

Looking Forward

Collection of the corpus has been completed for the 30 female talkers and our initial impression based on listening to the recordings is that we have captured some of the dialect variation we were looking for, despite the somewhat unnatural nature of the read speech materials. We expect to begin using these speech samples in the coming months in a number of novel perceptual and acoustic studies of the variation and variability of the English spoken in the state of Indiana.

In collecting this large corpus of speech under controlled laboratory conditions, we have gained valuable knowledge and insights into the kinds of read materials that are more or less likely to elicit variability in talkers from different regions. We are now working to streamline the recording process and to reduce the materials in order to make the project more feasible for extension to other parts of the country. In particular, we expect to include only half the number of sentences and multisyllabic words, a greatly reduced number of CVC words, and extended passage and spontaneous speech tasks in our larger study, the Nationwide Speech Project (NSP). The NSP will include recordings of talkers in Boston, New York City, Philadelphia, Atlanta, Chicago, Minneapolis, Dallas, Boulder, Los Angeles, and Seattle.

References

Baker, K., Ramig, L., Johnson, A., & Freed, C. (1997). Preliminary voice and speech analysis following fetal dopamine transplants in 5 individuals with Parkinson disease. *Journal of Speech, Language, and Hearing Research*, *40*, 615-626.

Bergquist, J.M. (1981). Tracing the origins of a midwestern culture: The case of central Indiana. *Indiana Magazine of History,* 1-32.

Byrd, D. (1992). Sex, dialects, and reduction. *ICSLP 92 Proceedings*, *92*, 827-830.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15,* 39-54.

Carmony, M. (1965). Aspects of regional speech in Indiana. In L.M. Davis (Ed.), *Studies in Linguistics in Honor of Raven I. McDavid, Jr.* (pp. 9-24). Tuscaloosa, AL: University of Alabama Press.

Carter, A.K. & Clopper, C.G. (2000). Prosodic and morphological effects on word reduction in adults: A first report. In *Research on Spoken Language Processing Progress Report No. 24* (pp. 67-86). Bloomington, IN: Speech Research Laboratory, Indiana University.

Carver, C.M. (1987). *American Regional Dialects: A Word Geography.* Ann Arbor, MI: University of Michigan Press.

Cassidy, F. (1985). *Dictionary of American Regional English.* Cambridge, MA: Belknap Press of Harvard University Press.

Clopper, C.G. (2000). Some acoustic cues for categorizing American English regional dialects: An initial report on dialect variation in production and perception. In *Research on Spoken Language Processing Progress Report No. 24* (pp. 43-66). Bloomington, IN: Speech Research Laboratory, Indiana University.

Clopper, C.G. & Pisoni, D.B. (2002). Some new experiments on perceptual categorization of dialect variation in American English: Acoustic analysis and linguistic experience. Invited lecture at the 9th Biennial Rice Symposium on Linguistics, Houston, TX. March 14-16.

Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor Speech Disorders*. Philadelphia: Saunders.

Davis, L.M. & Houck, C.L. (1992). Is there a Midland dialect? – Again. *American Speech, 67*, 61-70.

DuBois, J.W., Chafe, W.L., Meyer, C., & Thompson, S.A. (Eds). (2000). The Santa Barbara Corpus of Spoken American English. Santa Barbara, CA: University of California, Santa Barbara Center for the Study of Discourse.

Fairbanks, G. (1940). *Voice and Articulation Drillbook*. New York: Harper.

Fisher, W.M., Doddington, G.R., & Goudie-Marshall, K.M. (1986). The DARPA speech recognition research database: Specifications and status. *Proceedings of the DARPA Speech Recognition Workshop*, 93-99.

Gelfer, M. & Schofield, K. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice, 14*, 22-33.

Gerstman, L.H. (1968). Classification of self-normalized vowels. *IEEE Transactions of Audio and Electroacoustics AU-16*, 78-80.

Gibbens, V.E. (1962). Progress report on a word geography of Indiana. *Midwest Folklore, 11,* 151-154.

Hammond, M. (1999). *The Phonology of English*. Oxford: Oxford University Press.

Hayes, B. (1995). *Metrical Stress Theory*. Chicago: Chicago University Press.

Hillenbrand, J. & Houde, R. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39*, 311-321.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Karl, J.R. & Pisoni, D.B. (1994). Effects of stimulus variability on recall of spoken sentences. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193). Bloomington, IN: Speech Research Laboratory, Indiana University.

Keating, P., Blankenship, B., Byrd, D., Flemming, E., & Todaka, Y. (1992). Phonetic analyses of the TIMIT corpus of American English. *ICSLP 92 Proceedings*, *92*, 823-826.

Keating, P.A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*, *14*, 131-142.

Kelly, M. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27*, 343-358.

Kelly, M. & Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 389-403.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2,* 205-254.

Labov, W., Ash, S., & Boberg, C. (in press). *Atlas of North American English.* Mouton deGruyter.

Labov, W., Yeager, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress.* Philadelphia: US Regional Survey.

Linguistic Data Consortium. (1996). CallFriend Telephone Speech Corpus for American English. (LDC96S46 & LDC96S47).

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

McHenry, M. (1999). Aerodynamic, acoustic and perceptual measures of nasality following traumatic brain injury. *Brain-Injury, 13*, 281-290.

Milroy, J. & Milroy, L. (1993). Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics, 3*, 57-77.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Peterson, G. & Barney, H. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Sapienza, C., Walton, S., & Murry, T. (1999). Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *Journal of Speech, Language, and Hearing Research, 42*, 127-140.

Thomas, E.R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. Durham, NC: Duke University Press.

Torretta, G.M. (1995). The "easy-hard" word multi-talker speech database: An initial report. In *Research on Spoken Language Processing Progress Report No. 20* (pp. 321-334). Bloomington, IN: Speech Research Laboratory, Indiana University.

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, *9*, 351-356.

Appendix 1

Repeated CVC Words

bean	bite	boat	boil
bull	calm	can	cap
caught	coal	code	coin
con	cool	cot	cough
death	dig	dime	dock
doll	doubt	dull	fade
fail	feed	fell	fire
fool	foul	full	gap
good	guide	head	heal
hill	home	keep	lit
loud	love	loyal	luck
lull	lung	main	math
meal	mile	mill	mob
pal	pen	pin	poke
pool	pull	rice	rip
sail	sell	sour	south
tape	tool	towel	town
tube	voice	void	walk
wall	wet	wool	wrong

Appendix 2

Disyllabic Nonwords

Sw Words		wS Words	
defloing	pagrawl	bockis	linnid
degroff	pecluff	bozen	lofek
deploze	pefrowl	cheelut	moggit
deprinn	recleal	chueliff	noilup
detrath	reglann	doolis	shullis
detrull	rekress	fautchep	soggith
diprool	retroud	fiden	taylup
fegroff	sebrull	foutik	teelub
fegroish	seclide	guidet	thellig
medrall	seglutch	hanneb	tullib
mitween	seplail	kowtiff	villut
nebrenn	sigrine	koysig	vullek
nebruel	tebroze	kulleff	whelid
neprale	teprill	lannis	yailub

Appendix 3

The Rainbow Passage (Fairbanks, 1940)

When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for a pot of gold at the end of the rainbow.