

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 24 (2000)  
*Indiana University*

**A Multi-Talker Dialect Corpus of Spoken American English:  
An Initial Report on Development<sup>1</sup>**

**Cynthia G. Clopper, Allyson K. Carter, Caitlin M. Dillon, James D. Harnsberger,  
Rebecca Herman, Connie M. Clarke,<sup>2</sup> David B. Pisoni and Luis R. Hernandez**

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

---

<sup>1</sup> This work was supported by the NIH-NIDCD R01 Research Grant DC00111 and the NIH-NIDCD T32 Training Grant DC00012 to Indiana University. The authors would like to thank Julie Auger for her assistance in the design of the corpus.

<sup>2</sup> Psychology Department, University of Arizona, Tucson, Arizona

## **A Multi-Talker Dialect Corpus of Spoken American English: An Initial Report on Development**

**Abstract.** A multi-talker multi-dialect corpus of spoken American English has been designed to provide researchers who are interested in variation and variability with a large number of speech samples from twenty talkers in each of four cities located in phonologically distinct dialect regions of the United States: West (Los Angeles), South (Atlanta), Midland (Indianapolis), and Northern Cities (Chicago). The speech samples to be collected include word-length, sentence-length, and paragraph-length utterances, and have been designed to elicit phonological forms that differentiate the four regions. Once collected, these materials can be used for a range of perceptual and acoustic studies investigating the perception and production of dialect variation in the United States.

### **Objectives of the Corpus**

The purpose of this project is to create a speech corpus containing recordings from a large number of talkers from phonologically distinct dialect regions in the United States for use in a range of perceptual studies and acoustic analyses. Dialect variation, both regional and social in origin, has been an important topic of research in American English since the 1930's when plans for a "Linguistic Atlas of North America" were first discussed (Cassidy, 1993). The first studies were primarily concerned with regional variation, focusing on differences in lexical items produced by older males from rural areas (Chambers, 1993). More recently, dialect research has been extended to include studies on social and ethnic variation, such as African American Vernacular English and Appalachian English (Wolfram & Schilling-Estes, 1998).

Recent research has also begun to focus on phonological variation, particularly on variation and changes in progress that have been documented in the vowel systems of several American English dialects. For example, vowel shifts such as the Northern Cities Vowel Shift found in urban areas surrounding the Great Lakes and the Southern Vowel Shift found in rural areas of the Southern United States have been described in some detail (Wolfram & Schilling-Estes, 1998).

While such phonological variation has been studied via field recordings and transcription, relatively little work has been done to document the acoustic properties of these phenomena or to study their perceptual correlates via playback experiments. While acoustic analysis is a commonly accepted technique for comparing and differentiating the vowel systems of different languages, it is not commonly employed in sociolinguistic research due to Labov's "observer's paradox" (Wolfram & Schilling-Estes, 1998). Simply put, the paradox refers to the effect of the observer's presence (the observer being an experimenter, recording equipment, or any other tool of measurement) on the acoustic properties of speech produced by members of a dialect community of interest. The dialect variation that sociolinguists seek to document is almost always found in forms that appear in speech styles used more frequently in casual conversation, in specific pragmatic or situational contexts, or only with other members of the same dialect community. The intrusion of an experimenter from outside the dialect community and the effect of recording equipment on the formality of the conversational setting are perceived as barriers to the elicitation of the "deepest" form of the dialect in question (Wolfram & Schilling-Estes, 1998). Thus, the most commonly used method to investigate the properties of American English and other dialects involves making audio recordings of spontaneous speech and then phonetically transcribing those interviews.

While such methods are useful in describing relatively gross differences between dialects, they suffer from a number of limitations for researchers interested in the acoustic-phonetic properties of phonological forms of a dialect, and for researchers developing controlled stimulus materials varying in dialect for use in perception experiments. First, the use of spontaneous speech entails a lack of control over the particular stimulus materials elicited. For the experimenter hoping to collect numerous tokens of a particular vowel or word in a common phonetic and prosodic context, it is very difficult to elicit such materials in a natural, spontaneous speech style (cf. Harnsberger & Pisoni, 1999). While certain tasks, such as topically-guided conversations or map tasks, can be used to elicit particular words or prosodic phrases, strict control over the phonetic context of these forms cannot be achieved. Control of phonetic context is crucial for any acoustic analysis, as well as in constructing stimulus materials for use in perception tests.

Given these constraints, and given the purposes of this corpus, we have chosen to elicit speech materials in a read speech style, enabling control over the materials elicited. For the purposes of comparison only, we will also elicit a spontaneous sample from each talker, taking the form of a conversation with the experimenter administering the tests. While eliciting read speech undoubtedly limits the range of phonological variability we will observe between the dialects, we hope to ameliorate this problem by selecting American English dialects that have been shown in prior research to differ robustly from one another in terms of phonological patterns. We are also interested in documenting American English dialects that constitute relatively large communities within the United States. We believe that this will make the corpus as a whole more representative of American English dialectal variation than a corpus that is focused on much smaller dialect communities. We have therefore decided to record twenty talkers from each of four cities, representing four phonologically distinct regions: Atlanta (South), Indianapolis (Midland), Chicago (Northern Cities), and Los Angeles (West). For summary descriptions of each of the regional dialects, and for the rationale behind the selection of the boundaries defining these regions, see Wolfram and Schilling-Estes (1998) and Labov, Ash, and Boberg (1997). While we recognize that these four cities are not representative of all dialects of American English, we expect that they will provide us with some degree of phonological variation that is both acoustically and perceptually prominent, from a relatively large sample of talkers.

The nature of the controlled stimulus materials, the focus on dialect variation, and the large number of talkers we plan to record are the three main features that set this corpus apart from other existing corpora. There are at least three existing spoken language corpora that include speech samples from a large number of talkers from a variety of American English dialects: the Santa Barbara Corpus of Spoken American English (LDC Catalog, 2001c), the CALLFRIEND project (LDC Catalog, 2001a; LDC Catalog, 2001b), and the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990). The Santa Barbara corpus contains spontaneous speech samples from talkers from a wide range of geographic and socioeconomic backgrounds. The CALLFRIEND project contains recordings of telephone conversations between talkers which are grouped into two broad dialect categories: Southern and Non-Southern. The TIMIT Corpus contains ten read sentences from each of 630 talkers who come from eight defined dialect regions of the United States. The usefulness of the first two corpora in perceptual studies is limited by the lack of common stimulus materials for all talkers. The usefulness of the TIMIT corpus is also limited because of the ten sentences read by each talker, only two of those sentences were read by all 630 talkers.

Spoken language corpora that control for stimulus materials also exist. However, they do not necessarily vary the dialect of the talkers in a systematic fashion. For example, corpora used in our lab such as the “Easy-Hard” Word Multi-Talker Speech Database (Torretta, 1995) and the Talker Variability Sentence Database (Karl & Pisoni, 1994) contain fixed sets of stimuli spoken by 10-20 talkers, but no effort was made to identify or control for dialectal variation in the talkers. The new corpus will combine

the systematic variation in dialect found in the TIMIT corpus with the control over a range of stimulus materials found in the “Easy-Hard” and Talker Variability databases.

Once the corpus has been collected, we plan to use it in our lab for perceptual studies involving dialect identification, categorization, and discrimination by non-native listeners, lexical decision tasks, and voice quality judgement tasks involving dialect manipulations. This corpus will also be used in a series of perceptual learning tasks on dialect intelligibility after laboratory training and dialect manipulations in voice learning. Finally, the corpus will enable us to conduct acoustic-phonetic studies including descriptions of the vowel systems, analyses of diphthongal differences, and investigations into the acoustic correlates of stress across dialects.

## **Organization of the Corpus**

### **Talkers**

Ten males and ten females will be recorded in each of the four cities. Each talker will be a college-aged monolingual native speaker of English, with no history of hearing or speech disorders. In order to obtain a fairly homogenous group of talkers in terms of socioeconomic status, level of education, and linguistic experience, talkers will be recruited from community college campuses and will be asked to complete a lengthy questionnaire. In order to participate, a talker must have lived in the city of interest for his or her entire life and have limited experience with other dialects and languages. Parents of the talkers must also be native English speakers who are local to the area.

### **Stimulus Materials**

The materials list was selected to provide a number of different kinds of speech, including word-length, sentence-length, and paragraph-length materials. The materials themselves were selected with the intent of providing a useful corpus for completing the projects mentioned above.

The word-length materials include CVC’s and multisyllabic words and nonwords. The CVC list was designed for this project and consists of 1020 CVC’s selected from an online dictionary containing approximately 20,000 entries based on Webster’s Pocket Dictionary. The list is composed of all CVC’s in the dictionary that received a familiarity rating of 6.0 or greater (on a 7-point scale) by undergraduates (Nusbaum, Pisoni, & Davis, 1984). A small subset of these CVC’s was hand-selected for an additional repetition in recording. This subset was selected such that the vowels occurred in consonantal contexts that are expected to reveal systematic differences between the dialects, based on documented shifts and mergers (Callary, 1975; Gordon, 1997; Labov, 1972; Labov, Yeager, & Steiner, 1972; Wolfram & Schilling-Estes, 1998). Additionally, 10 vowels will be recorded in an “hVd” context for use in determining the vowel space of each talker (Hillenbrand, Getty, Clark, & Wheeler, 1995; Hagiwara, 1997). The multisyllabic word list is a subset of the list developed by Carter and Clopper (this volume), and contains 240 words that vary systematically in the number of syllables and the location of primary stress. The multisyllabic nonword list was developed for this project and contains 56 disyllabic forms. These forms have been designed so that half will be realized with primary stress on the first syllable and the other half will be realized with primary stress on the second syllable (Cutler & Carter, 1987; Hammond, 1999; Hayes, 1995; Kelly, 1988; Kelly & Bock, 1988).

The sentence-length materials include high probability, low probability and anomalous sentences. The high probability sentences were taken from all eight of the Speech Perception in Noise (SPIN) lists (Kalikow & Stevens, 1977), with several additional sentences taken from the Hearing in Noise Test (HINT) sentence list (Nilsson, Soli, & Sullivan, 1994) to round out the representation of all English

vowels in the content words in the sentences. In high probability sentences, the final target word is predictable based on the preceding words in the sentence. In low probability sentences, the final target word is not predictable from the rest of the sentence. The low probability sentences were taken from lists 1, 2, 7, and 8 of the SPIN test. The anomalous sentences were created from the SPIN sentences, so that their target words matched those for the low probability sentences that were selected. The remaining words were taken from the high probability sentences in the remaining four lists, using a method similar to Miller and Isard (1963).

The longer materials include a passage and a spontaneous speech sample. The passage selected was the Rainbow Passage (Fairbanks, 1940). This passage has a long history of use in perceptual and acoustic studies, including several involving individual differences and variability (Gelfer & Schofield, 2000; Sapienza, Walton, & Murry, 1999). The short spontaneous speech sample will focus mainly on discussions about the local geographic area and will be used primarily as a reference point for each talker.

## Methods

### Recording

All recording will be done in sound-attenuated booths located in each city. Materials will be presented visually to the talkers via a portable Macintosh Powerbook G3 computer and the talkers will be asked to read the materials aloud into a head-mounted dynamic unidirectional cardioid microphone (Shure SM10A) as they are presented. Responses will be recorded digitally in real time into individual sound files on the computer and simultaneously on DAT, using a Sony TC8 recorder, as a backup. The nine stimulus sets will be presented in a pseudo-random order, and all stimuli within each set will also be presented randomly.

### Future Directions

Collection of the data is expected to begin in the Spring of 2001. We hope to complete the data collection within six months and to have all of the speech available on CD-ROM, with documentation shortly thereafter.

## References

- Callary, R. (1975). Phonological change and the development of an urban dialect in Illinois. *Language in Society*, 4, 155-169.
- Carter, A.K. & Clopper, C.G. (this volume). Prosodic and morphological effects on word reduction in adults: A first report.
- Cassidy, F.G. (1993). Area lexicon: the making of DARE. In *American Dialect Research*. Preston, D.R. (ed.) Philadelphia, PA: John Benjamins, 93-106.
- Chambers, J.K. (1993). Sociolinguistic dialectology. In *American Dialect Research*. Preston, D.R. (ed.) Philadelphia, PA: John Benjamins, 133-164.
- Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- Fairbanks, G. (1940). *Voice and Articulation Drillbook*. New York: Harper.
- 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.
- Gordon, M.J. (1997). Urban sound change beyond city limits: The spread of the northern cities shift in Michigan. Doctoral dissertation, The University of Michigan.

- Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102, 655-658.
- Hammond, M. (1999). *The Phonology of English*. Oxford: Oxford University Press.
- Harnsberger, J.D. & Pisoni, D.B. (1999). Eliciting speech reduction in the laboratory II: Calibrating cognitive loads for individual talkers. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 339-349). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Hayes, B. (1995). *Metrical Stress Theory*. Chicago, IL: Chicago University Press.
- Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.
- Kalikow, D.N. & Stevens, K.N. (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: A first report. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193). Bloomington, IN: Speech Research Laboratory, Indiana University.
- 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. (1972). The internal evolution of linguistic rules. In *Linguistic Change and Generative Theory*. Stockwell, R.P. & Macaulay, R.K.S. (eds.) Bloomington, IN: Indiana University Press, 101-171.
- Labov, W., Ash, S., & Boberg, C. (1997). A National Map of the Regional Dialects of American English. Retrieved June 26, 2000 from the World Wide Web: [http://www.ling.upenn.edu/phono\\_atlas/NationalMap/NationalMap.html](http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html).
- Labov, W., Yeager, M., & Steiner, R. (1972). A quantitative study of sound change in progress. Philadelphia, PA: U.S. Regional Survey.
- LDC Catalog. (2001a). CALLFRIEND American English Non-Southern Dialect. Retrieved January 9, 2001 from the World Wide Web: <http://www ldc.upenn.edu/Catalog/LDC96S46.html>.
- LDC Catalog. (2001b). CALLFRIEND American English Southern Dialect. Retrieved January 9, 2001 from the World Wide Web: <http://www ldc.upenn.edu/Catalog/LDC96S47.html>.
- LDC Catalog. (2001c). Santa Barbara Corpus of Spoken American English Part-I. Retrieved January 9, 2001 from the World Wide Web: <http://www ldc.upenn.edu/Catalog/LDC2000S85.html>.
- Miller, G.A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2, 217-228.
- Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.
- 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.
- 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.
- 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.
- Wolfram, W. & Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell.
- Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, 9, 351-356.