The “Up” corpus: A corpus of speech samples across adulthood

Susanne Gahl, Emily Cibelli, Kathleen Hall, Ronald Sprouse

This is an Accepted Manuscript of an article in press in Corpus Linguistics and Linguistic Theory. The article is available online at:

https://www.degruyter.com/view/j/cllt.2014.10.issue-2/cllt-2013-0023/cllt-2013-0023.xml

Permanent link: http://DOI 10.1515/cllt-2013-0023

The “Up” corpus: A corpus of speech samples across adulthood

Susanne Gahl, Emily Cibelli, Kathleen Hall, Ronald Sprouse

Affiliation:

Linguistics Department
University of California at Berkeley
1203 Dwinelle Hall,
Berkeley, California 94720

Contact information:

gahl@berkeley.edu
ecibelli@berkeley.edu
hall.kat.g@gmail.com
ronald@berkeley.edu

Running title: “Speaking Up”: A longitudinal corpus

We describe a speech corpus based on the “Up” series of documentary films by director Michael Apted, showing a set of individuals at seven year intervals over a period of 42 years. The corpus is meant to facilitate phonetic, psycholinguistic and sociolinguistic research on age-related change in speech during young and middle-age adulthood. The corpus contains audio files, transcripts time-aligned at the level of utterance, word, and segment, F0 and vowel formant measurements of portions of the films featuring eleven participants at ages 21 through 49. The corpus is freely available to researchers upon request.

Key words: aging, vocal aging, vowel formants, talker variability, Sociolinguistics, talker age, speaking tempo, language development, corpora, Michael Apted (director).

1. Introduction

Many theories of adult language processing make the simplifying assumption that the language processing system reaches a stable state after puberty and remains stable until old age, at which point it becomes subject to age-related decline fairly suddenly. And yet, converging research programs in Phonetics, Sociolinguistics, and Psycholinguistics (both experimental and corpus-based) have documented age-related changes in speech and language processing (Endres et al. 1971; Horton et al. 2010; Lima et al. 1991; Mortensen et al. 2006; Mysak and Hanley 1959; Ramig 1986; Ramig et al. 2001; Ramig and Ringel 1983; Trudgill 1988). Investigations of age-related changes have often proceeded by comparing groups of young and old speakers (e.g. Duchin and Mysak 1987; Harnsberger et al. 2008). On the one hand, such cross-sectional studies offer some important information not apparent in studies tracking individuals over time (cf. Braun and Friebis 2009, for discussion). On the other hand, cross-sectional studies suffer from limitations: For example, cross-sectional demographic data of many parts of the world show differences in ethnic composition of different age groups, due to migration and other demographic shifts. Such data would be consistent with the erroneous conclusion that people's ethnicity changes as they age. More generally, cross-sectional samples cannot differentiate individual development from static differences across groups.

Cross-sectional studies are complemented by longitudinal studies, i.e. studies following individuals through time. Several studies along these lines have proceeded by re-contacting middle-aged or elderly groups of speakers first documented in sociolinguistic interviews during their young adult or middle-aged years. Examples of such long-term follow-up studies include Sankoff and Blondeau (2007), Trudgill (1988), and Mather (2012). Another resource based in part on multiple samples from the same talkers over long time periods is the ONZE project on the development of New Zealand English (Gordon et al. 2007). A major longitudinal study of written language is the 'Nun Study' (Snowdon et al. 1996), an analysis of autobiographic writing samples produced by a group of nuns when they entered the convent (between 18 and 32 years of age) and about sixty years later (75 to 93 years of age). A number of studies (e.g. Sancier and Fowler 1997) document individual changes in pronunciation over fairly short periods of time (weeks or months, rather than years or decades). These resources have yielded a great deal of information, yet they are typically limited to probing two or three points in any participant's life.

Other studies have traced changes in the speech of famous individuals over time through greater numbers of samples. One class of studies along these lines has focused on the writings of authors whose literary works span several decades, such as Iris Murdoch, Agatha Christie, and P.D. James (Le et al. 2011) and the twenty-two letter writers (most of them famous authors) examined in Arnaud (1998). Analyses of vocal aging have similarly focused on figures of public life, including the Queen of England (Harrington 2006; Harrington et al. 2000), former British Prime Minister Margaret Thatcher and the British-American journalist Alistair Cooke (Reubold et al. 2010). Observations of public behavior by public figures are not only easily obtained, in the United States they are also exempt from federal policy regulating human subjects research, known as the "Common Rule" for Protection of Human Subjects (United States Department of Health and Human Services 1991, Revised 2009: 45 CFR, Part 46, Protection of Human Subjects). However, a limitation of these studies is that they are by necessity based on carefully considered spoken remarks or on written language, rather than on spontaneous speech.

Longitudinal studies of acoustic characteristics of speech tracking the spontaneous speech of individuals through multiple samples continue to be rare. In part, this is due to difficulties inherent in the collection of longitudinal data in real time, accompanied as it is by the researcher's own aging process. An additional difficulty is that the mere availability of longitudinal data is not enough. Archives of long-running soap operas, long-term TV and radio personalities, and so on, provide longitudinal speech samples of essentially unlimited size, after all – but absent transcription, segmentation, and other steps in preprocessing, such resources are of limited usefulness.

The current paper describes a project aimed at helping to remedy that scarcity of accessible pre-processed data. The corpus is based on the "Up" series of documentary films by director Michael Apted, showing a set of individuals at seven year intervals over a period of 42 years (Apted 1977, 1984, 1991, 1998). The documentary was filmed in the UK, and the participants grew up in various parts of the UK, including London, Yorkshire, and Liverpool.

The ethical and legal considerations involved in studying private individuals are complicated and need to be considered carefully. When the participants were first filmed, at age 7, they would not have been in a position to give informed consent even if such consent had been sought at the time. The fact that participants were enrolled in the study without their informed consent is one

reason we refrained from analyzing speech samples from ages 7 or 14. Several participants declined to participate in the films as adults (starting at age 28), and we refrained from including any data from these participants in the database. By the time the participants had reached age 28, the enormous success and popularity of the films had become clear; therefore, participants who, as adults, agreed to continue their participation knew at that point that their participation in a documentary film would not be private behavior. A successful documentary film series does not constitute a context in which a participant can have a "reasonable expectation of privacy", as federal guidelines put it. Legally, the interview materials would thus appear to be on a similar footing as speech and other data from figures of public life, such as radio broadcasts. Ethically, analyzing the films strikes us as problematic in some respects. Some participants register frustration, during the later films, with the effects the films have had on their lives. For the current project, data from these participants were therefore excluded from analysis.

Another relevant legal consideration concerns copyright. Since the likely uses of the corpus are not of a commercial nature, since amount of data analyzed is tiny compared to the total length of the movies, and since we consider it highly unlikely that users will choose to read the orthographic transcript of the database in lieu of watching the movies, we consider that the current work falls under the "fair use" doctrine under United States copyright law.

A number of previous studies have already made use of the Up series for sociolinguistic research. The first study to do so (Sankoff 2004) focused on two sets of vowels ([a, ɑ] and [ʊ/ʌ]) in the speech of two individuals at ages seven through thirty-five. The analysis of these vowel pairs was based on categorizing vowel tokens into three broad classes (e.g. [a] vs. [ɑ] vs. "intermediate" between these two) by a single transcriber. One possibly problematic aspect of that procedure is that vowels are notoriously difficult to categorize based on auditory impression alone. For example, one study of a corpus of spontaneous speech (Pitt et al. 2005) reports agreement of 64% for [a], 67% for [ɑ], 55% for [ʊ], and 48% for [ʌ]. Coded speech samples are therefore maximally useful if accompanied by acoustic data and analysis. Two subsequent studies to make use of the Up series (Hansen 2007; Poplack and Lealess 2009) focused on two additional linguistic variables (h-dropping and theta-fronting) in 11 individuals. Both studies underscore the usefulness of the films as linguistic data, but neither resulted in publicly available annotated speech corpora.

To address the need for more data, including acoustic measurements, the current study used automatic time-alignment of orthographic transcripts and audio and automatic formant extraction. We provide a sizable amount of material ready for further analysis. Access to the data is in the form of a website making available the orthographic transcripts, formant measurements, and audio files.

2. The corpus

The speech samples are taken from five films from the film series known as the "Up" series of documentary films (Apted 1977, 1984, 1991, 1998). The films follow fourteen individuals, first filmed when they were seven years old, and again every seven years. The first film, featuring the participants at age seven, was released in 1964. The most recent film included in our database shows the participants at age 49. Ten of these individuals have participated in all films released so far.

We measured fundamental frequency (F0), the first four vowel formants, and speaking rate. Wherever possible, we focused on utterances of at least 30 seconds or more of uninterrupted speech (using the absence of interruptions, or pauses longer than 500 ms, as the criterion of what to consider as an "utterance" for the purposes of annotation), seeking to analyze at least one such utterance from each talker at each age. In all cases, utterances selected for analysis were at least 20 seconds in duration.

Each utterance selected for analysis is annotated for the film it appeared in, the section within the film, and exact start and end times. The time stamps are included to allow future users of the database to link audio and video, to facilitate the analysis of visual information such as facial expressions, gaze direction, gestures. Once utterances were selected and coded for talker, age, and film, transcribers produced an orthographic transcript.

The audio files were aligned with the transcripts at the phone level using the Penn Phonetics Lab Forced Aligner Toolkit (Yuan and Liberman 2008). The aligned files were hand-checked by a research assistant with training in Phonetics, and the alignment procedure was repeated in cases where the alignment failed. In most cases, such failures turned out to be due to errors or omissions in the orthographic transcript. Additional misalignments were corrected by dividing

misaligned utterances into smaller segments and repeating the alignment procedure on each, a strategy which ensured tighter correspondence between the orthographic transcription and the speech sample. It should be noted that the Penn Forced Aligner is built on acoustic models of American English, while our sample is drawn from varieties of British English. However, the resulting differences in labeling did not interfere with the analysis. Inspection of the results revealed that these differences did not affect the ability of the aligner to align the transcript with the signal. Critically, use of the aligner provides access to a much larger set of utterances than we could analyze if the data were transcribed by hand.

The start and end times of each vowel phone was obtained from the alignment results, and a portion of each token's audio file was extracted, starting 40ms before the start time and ending 40ms after the end time of the vowel. This audio was downsampled to 12 KHz and analyzed by the Watanabe and Ueda formant tracker (Ueda et al. 2007).

The corpus is segmented and labeled at the utterance, word and phone level. “Word” were defined as strings of letters separated by white space in the orthographic transcript. Words thus include contracted forms. Figure 1 shows a sample utterance (“Give me the boy until he is seven”) at the annotation stage.

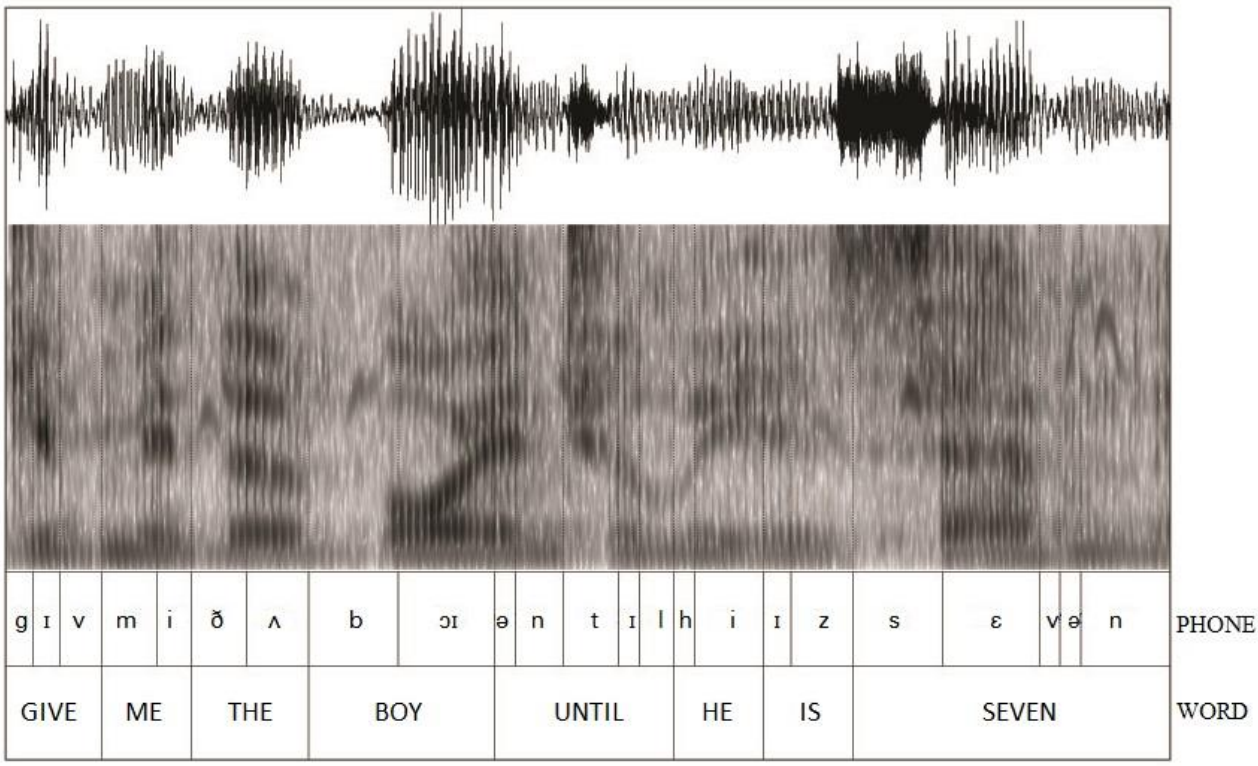

Figure 1: Annotation of part of an utterance ("Give me the boy until he is seven"), including waveform, spectrogram, segmentation into phones, and orthographic transcript.

The corpus includes plain text files with the orthographic and phonemic transcriptions, start and end time stamps of each word and phone, and formant measurements (in the case of vowels). Figure 2 shows a part of the file with the phone-level information corresponding to the speech segment in Figure 1. Phonetic symbols are rendered as IPA symbols here, but are in ACII-readable characters in the database.

ptoken_id	phone	ptoken_start	ptoken_end	word	speaker	age	sex
346674	g	624.44	624.48	GIVE	nick	35	male
346675	ɪ	624.48	624.52	GIVE	nick	35	male
346676	v	624.52	624.58	GIVE	nick	35	male
346677	m	624.58	624.66	ME	nick	35	male
346678	i	624.66	624.71	ME	nick	35	male
346679	ð	624.71	624.79	THE	nick	35	male
346680	ʌ	624.79	624.88	THE	nick	35	male
346681	b	624.88	625.01	BOY	nick	35	male
346682	ɔɪ	625.01	625.15	BOY	nick	35	male
346683	ə	625.15	625.18	UNTIL	nick	35	male
346684	n	625.18	625.25	UNTIL	nick	35	male
346685	t	625.25	625.33	UNTIL	nick	35	male
346686	ɪ	625.33	625.36	UNTIL	nick	35	male
346687	l	625.36	625.41	UNTIL	nick	35	male
346688	h	625.41	625.44	HE	nick	35	male
346689	i	625.44	625.54	HE	nick	35	male
346690	ɪ	625.54	625.58	IS	nick	35	male
346691	z	625.58	625.67	IS	nick	35	male
346692	s	625.67	625.8	SEVEN	nick	35	male
346693	ɛ	625.8	625.94	SEVEN	nick	35	male
346694	v	625.94	625.97	SEVEN	nick	35	male
346695	ə	625.97	626	SEVEN	nick	35	male
346696	n	626	626.25	SEVEN	nick	35	male

Figure 2: Sample file with IPA symbol, start and end times of each phone, and orthographic transcription of part of an utterance (“Give me the boy until he is seven”).

The corpus comprises 250 utterances (21,328 word tokens) produced by eleven of the documentary participants, yielding 27,921 vowel tokens and 41,284 consonant tokens. For nine of the participants (Andrew, Bruce, John, Lynn, Neil, Nick, Paul, Sue, Suzi, and Tony), we were able to find utterances for every age that met our inclusion criteria. Table 1 shows the number of utterances, words, and syllables analyzed for each talker, as well as the number of word tokens analyzed for each of the talkers at each age.

INSERT TABLE 1 ABOUT HERE

The 21,328 word tokens represent 2463 unique word types, including 10,944 tokens (2322 types) of open-class words ("content words") and 10,384 tokens (141 types) of closed-class words ("function words"). Function words were defined as pronouns (e.g. *I, somebody*), determiners (e.g. *the, many*), complementizers (e.g. *whether*), and conjunctions (e.g. *albeit, because*), as well as contracted forms such as *they've, they're, that's*, which are treated as single word units by the aligner. The database contains a total of 69,205 phones (41,284 consonants and 27,921 vowels). Table 2 shows the number of tokens for each phone, for each age.

INSERT TABLE 2 ABOUT HERE;

As an example of the kinds of information that can be easily extracted from the corpus, we include in Figure 2 a view of the vowel space, based on the first two formants, of one talker (Nick) from age 21 through 49, at the analysis frame occurring at the temporal midpoint of the vowel. The graph is based on all tokens of vowels with primary stress in monosyllabic content words.

Vowel formants vary as a function of many factors, such as speaking rate and phonological context. Tracking individual talker's vowel spaces, e.g. the space defined by the F1/F2 coordinates of vowel tokens, provides insights into speech production and perception (Bradlow et al. 1996; Byrd and Saltzman 2002; Ferguson 2007; Johnson et al. 1993; Lindblom 1963, 1964). Age-related change in vowel spaces has variously been reported to take the form of vowel centralization, i.e. overall contraction of the space, as well as a shift of formant peaks to lower frequencies (Endres et al. 1971; Harrington 2006). However, this tendency is not observed consistently across studies, suggesting that the effect is not uniform across talkers, and possibly suggesting an increase in variability of formant peaks with increasing talker age.

One interpretation of the overall pattern for the talker whose vowels are plotted in Figure 3 is that, with increasing talker age, the vowel space contracts, with vowel tokens more likely to be produced near the center of the talker's vowel space. In a separate study, we analyze the talkers' vowel spaces more closely, taking into account word-specific information, such as lexical frequency, that is known to affect the realization of vowels (Bell et al. 2003; Bybee 2001; Gahl et al. 2012). The plain-text format of the database makes it straightforward to combine it with information from other resources, such as information about lexical properties (e.g. Balota et al. 2007).

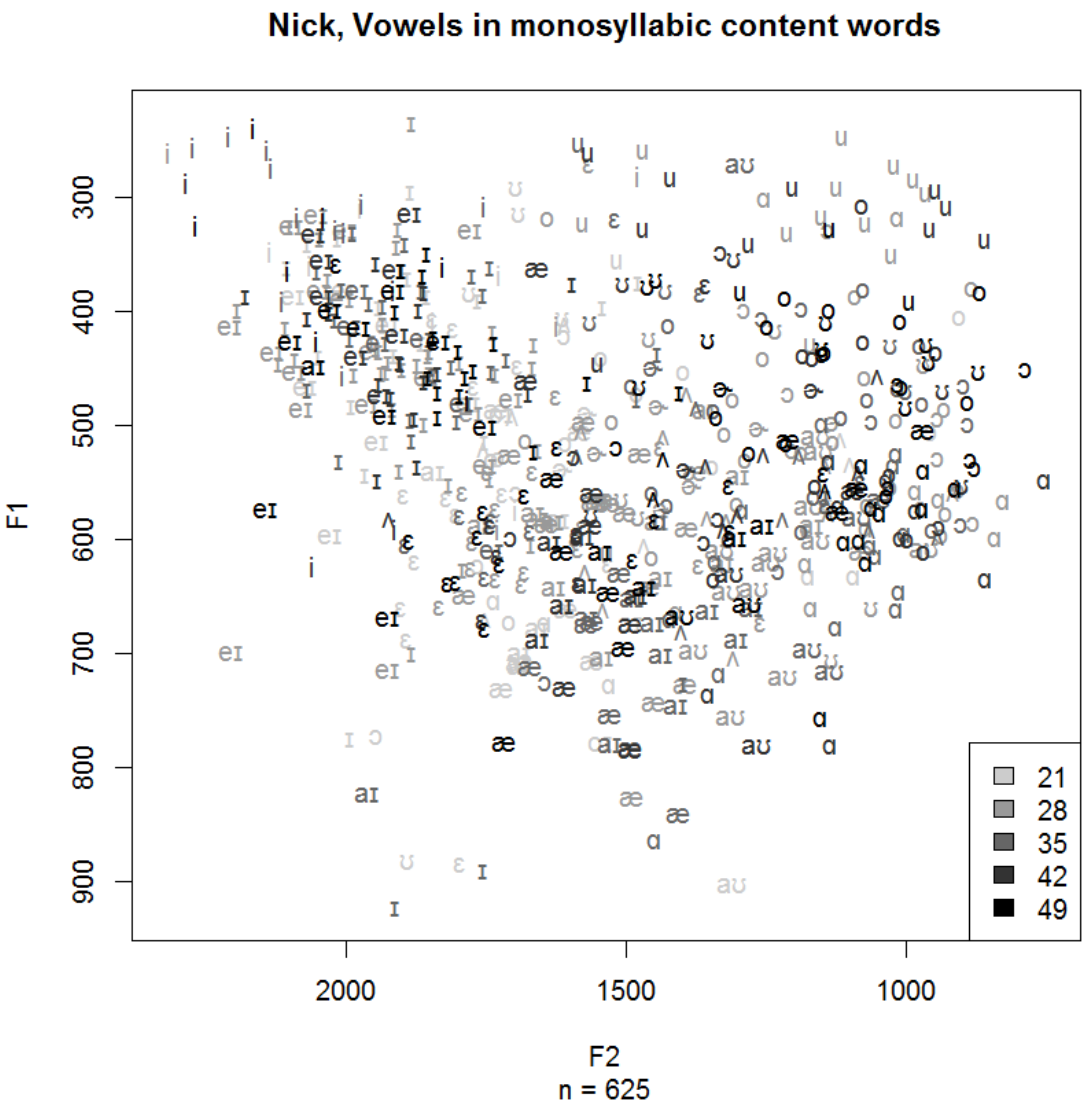

Figure 3: Formant values (F1 and F2) for vowels in monosyllabic content words in one talker, age 21 through 49.

3. Summary and conclusion

The "Up" database provides a comprehensive source of utterances recorded at regular intervals from young adulthood through middle age. The small, but growing literature of longitudinal analyses of speech in adulthood (e.g. Braun and Friebis 2009, Reubold et al. 2010) has so far mostly focused on the speech of public figures and has offered little opportunity to study spontaneous speech of private individuals, i.e. individuals who are not figures of public life. The resource described here seeks to fill that gap.

Like all corpora based on speech samples from small numbers of individual speakers, the "Up" Corpus has limitations. It is our hope that the corpus will grow, perhaps eventually including the interviewer's speech (who can be heard in all of the films). The substantial individual variability underscores the need for complementing such studies with large-scale studies and with controlled experiments before any definitive explanations as to the sources of individual variability can be attempted. For example, an observed change in vowel characteristics may arise due to an individual's move from one dialect region to another (Sankoff 2004), to changes in stylistic practice (Eckert 2008), to changes in language processing mechanisms (Mortensen et al. 2006), or to age-related physiological change in the vocal apparatus (Linville 2001). One necessary step towards an understanding of the role of physiological, social, cognitive, and linguistic factors in such changes are analyses showing what changes reliably co-occur in individuals and groups. It is our hope that the creation of this corpus will enable such research.

Research questions in Phonetics, language development, psycholinguistics, and sociolinguistics often need to be addressed against the backdrop of normal age-related change in speech: For example, an observed change in vowel formants might point to language change in progress, or to changes typically associated with vocal aging in middle-age adulthood. Archival recordings of unscripted conversational speech have the potential to aid such research**.** The "Up" series of documentary films provides one such source. Future versions of the database may also make use of the video material, for example for the study of events accompanying pauses and disfluencies in speech, including gestures, eye gaze, and interaction in discourse.

References

Apted, Michael. 1977. 21 Up: ITV Granada Television.

Apted, Michael. 1984. 28 Up: ITV Granada Television.

Apted, Michael. 1991. 35 Up: ITV Granada Television.

Apted, Michael. 1998. 42 Up: BBC.

Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10(2). 123-152.

Balota, David A., Melvin J. Yap, Michael. J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. The English Lexicon project. *Behavior Research Methods* 39. 445-459.

Bell, Alan, Dan Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Dan Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001-1024.

Bradlow, Ann R., Gina Torretta David B. & Pisoni. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20. 255-272.

Braun, Angelika & Stefan Friebis. 2009. Phonetic cues to speaker age: A longitudinal study. In Günter Grewendorf & Monika Rathert (eds.), *Formal Linguistics and Law*, 141-162. Berlin & New York: Mouton deGruyter.

Bybee, Joan 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Byrd, Dani & Elliot Saltzman. 2002. Speech production. In Michael A. Arbib (ed.), *The handbook of brain theory and neural networks*, 1072-1076. Cambridge: MIT Press.

Duchin, Sandra W. & Edward D. Mysak. 1987. Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of communication disorders* 20(3). 245-257.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12. 453-476.

Endres, Werner, W. Bambach & G. Floesser. 1971. Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America* 49(6B). 1842-1848.

Ferguson, Sarah H. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research* 50. 1241-1255.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4). 789-806.

Gordon, Elizabeth, Margaret Maclagan & Jennifer B. Hay. 2007. The ONZE Corpus. In Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora. Volume 2: Diachronic Databases*, 82-104. Basingstoke: Palgrave Macmillan.

Hansen, Kathrin. 2007. *H-Dropping as an indicator of independent social variables*. Potsdam: University of Potsdam MA thesis.

Harnsberger, James. D., Rahul Shrivastav, W. S. Brown, Jr, Howard Rothman & Harry Hollien. 2008. Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice* 22(1). 58-69.

Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34. 439-457.

Harrington, Jonathan, Sallyanne Palethorpe & Catherine Watson. 2000. Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetics Association* 30(1/2). 63-78.

Horton, William S., Daniel H. Spieler & Elizabeth Shriberg. 2010. A Corpus Analysis of Patterns of Age-Related Change in Conversational Speech. *Psychology and Aging* 25(3). 708-713.

Johnson, Keith, Edward Flemming & Richard Wright. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69. 505-528.

Le, Xuan, Ian Lancashire, Graeme Hirst & Regina Jokel. 2011. Longitudinal Detection of Dementia through Lexical and Syntactic Changes in Writing: A Case Study of Three British Novelists. *Literary and Linguistic Computing* 26(4). 435-461.

Lima, Susan D., Sandra Halea & Joel Myerson. 1991. How general is general slowing? Evidence from the lexical domain. *Psychology and aging 6*(3). 416-425.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773-1781.

Lindblom, Björn. 1964. Articulatory activity in vowels. *Speech, Music and Hearing, Quarterly Progress Report* 1. 1-15.

Linville, Sue Ellen. 2001. *Vocal Aging.* San Diego: Singular.

Mather, Patrick-André. 2012. The Social Stratification of /r/ in New York City: Labov's Department Store Study Revisited. *Journal of English Linguistics* 40(4). 338-356.

Mortensen, Linda, Antje S. Meyer & Glyn W. Humphreys. 2006. Age-related effects on speech production: A review [Special issue on language production across the life span]. *Language and Cognitive Processes* 21(1-3). 238-290.

Mysak, Edward D. & T. D. Hanley. 1959. Vocal aging *Geriatrics* 14. 652-656.

Pitt, Mark A., Keith Johnson, Elizabeth Hume, Scott Kiesling & William Raymond. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45. 89-95.

Poplack, Seana & Allison V. Lealess. 2009. *Further insights from the "Up" series*. Paper presented at the New Ways of Analyzing Variation conference, University of Ottawa.

Ramig, Lorraine A. 1986. Aging speech: Physiological and sociological aspects. [Special Issue: Language, communication and the elderly]. *Language & Communication* 6(1-2). 25-34.

Ramig, Lorraine A., Steven Gray, Kristin Baker, Kim Corbin-Lewis, Eugene Buder, Erich Luschei, Hillary Coon & Marshall Smith. 2001. The aging voice: A review, treatment data and familial and genetic perspectives. *Folia Phoniatrica et Logopaedica* 5. 252-263.

Ramig, Lorraine A. & Robert L. Ringel. 1983. Effects of physiological aging in selected acoustic characteristics of voice. *Journal of Speech & Hearing Research* 26(1). 22-30.

Reubold, Ulrich, Jonathan Harrington & Felicitas Kleber. 2010. Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication* 52. 638-651.

Sancier, Michele L. & Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25. 421-436.

Sankoff, Gillian. 2004. Adolescents, young adults, and the critical period: Two case studies from Seven Up. In Carmen Fought (ed.), *Sociolinguistic Variation: Critical Reflections,* 121-139. Oxford: Oxford University Press.

Sankoff, Gillian & Hélène Blondeau. 2007. Language Change across the Lifespan: /r/ in Montreal French. *Language* 83(3). 560-588.

Snowdon, David A., Susan J. Kemper, James A. Mortimer, Lydia H. Greiner, David R. Wekstein & William R. Markesbery. 1996. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association* 275(7). 528-532.

Trudgill, Peter. 1988. Norwich revisited: recent linguistic changes in an English urban dialect. *English World-Wide* 9. 33-49.

Ueda, Yuichi, Tomoya Hamakawa, Tadashi Sakata, Syota Hario & Akira Watanabe. 2007. A real-time formant tracker based on the inverse filter control method. *Acoustical Science and Technology of the Acoustical Science of Japan* 28(4). 271-274.

United States Department of Health and Human Services. 1991 [Revised 2009]. 45 CFR, Part 46, Protection of Human Subjects.

Yuan, Jiahong & Mark Liberman. 2008. *Speaker identification on the SCOTUS corpus. Proceedings of Acoustics* 08. 5687-5690.

Table 1: Number of utterances and words analyzed for each talker in the corpus. Cells marked "n/a" indicate films in which particular talkers declined to participate.

Talker	Number of utterances analyzed	Total number of word tokens analyzed	Number of word tokens at each age (Numbers in parentheses indicate the number of content words out of the total word count for each cell.)				
			21	28	35	42	49
Andrew	15	791	55 (31)	139 (75)	218 (132)	306 (173)	73 (36)
Bruce	30	2842	676 (349)	433 (228)	774 (415)	696 (367)	263 (132)
John	14	1662	538 (295)	n/a	694 (344)	n/a	430 (234)
Lynn	24	1222	277 (137)	140 (77)	213 (107)	275 (152)	317 (156)
Neil	34	3978	979 (464)	1010 (501)	982 (494)	480 (247)	527 (258)
Nick	28	2543	363 (190)	803 (446)	460 (241)	461 (248)	456 (225)
Paul	23	1653	397 (177)	304 (164)	219 (107)	279 (150)	454 (225)
Sue	18	1467	151 (74)	276 (139)	277 (150)	278 (147)	485 (238)
Suzi	24	2350	303 (152)	620 (322)	533 (275)	505 (263)	389 (199)
Symon	23	1318	548 (279)	426 (213)	n/a	344 (176)	n/a
Tony	17	1502	452 (222)	192 (98)	93 (44)	489 (244)	276 (132)
TOTAL	250	21,328					

Table 2: Number of tokens for each phone type at each talker age

Phone	Number of tokens produced at age ….				
	21	28	35	42	49
ɑ	212	221	227	208	178
æ	318	330	376	318	293
aɪ	581	470	487	468	400
aʊ	89	88	79	88	89
b	293	306	255	282	245
ɔ	211	139	169	182	143
ð	377	399	416	312	338
ʤ	84	75	67	89	71
d	679	592	624	568	547
ɛ	444	444	397	353	311
ɚ	309	307	327	285	249
eɪ	218	208	170	188	163
f	247	238	219	208	192
g	171	155	131	123	141
h	166	184	213	229	197
ɪ	804	730	813	651	657
i	561	543	591	557	536
j	151	168	143	138	106
k	405	397	385	371	320
l	560	536	577	481	451
m	488	447	492	463	390
n	1036	974	970	948	828
ŋ	255	223	247	177	164
o	223	221	194	245	169
oɪ	19	13	10	7	9
p	267	272	308	227	214
r	531	463	523	460	419
ʃ	59	59	89	85	77
s	695	587	655	558	491
ʧ	79	78	73	73	67

t	1167	1170	1149	1081	975
ʊ	90	78	75	65	51
u	261	241	244	238	183
ʌ	1769	1732	1718	1514	1372
v	299	294	315	300	244
w	404	379	403	299	348
z	337	329	364	329	298
ʒ	3	10	3	7	8
θ	160	142	144	97	93