Title
Speaking "Up": A Longitudinal Corpus Based on the "Up" Film Series

Journal
UC Berkeley PhonLab Annual Report, 8(8)

ISSN
2768-5047

Authors
Gahl, Susanne
Hall, Kathleen
Cibelli, Emily
et al.

Publication Date
2012

DOI
10.5070/P79176n6vg

Speaking "Up": A longitudinal corpus based on the "Up" film series

Susanne Gahl, Kathleen Hall, Emily Cibelli, Ronald Sprouse

Linguistics Department
University of California at Berkeley
Berkeley, California 94720

gahl@berkeley.edu

Abstract

This report describes 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 broader goal of the research reported here is to analyze age-related change in speech across the adult life span. The corpus is based on portions of the films featuring the participants at ages 21 through 49. The current report describes the resulting database, which is freely available upon request to researchers, as well as some sample analyses of speaking rate and vowel formants. These analyses reveal subtle changes in each talker's speaking rate and vowel spaces in the same direction as in previous studies investigating vocal aging, despite the much younger age range represented in this corpus.

I INTRODUCTION

Many theories of adult language processing make the useful simplifying assumption that the language processing system reaches a stable state after puberty and remains stable until the onset of age-related pathology. And yet, converging research programs in Phonetics, Sociolinguistics, and

Psycholinguistics have documented age-related changes in speech and language processing in young and middle-aged adults (see e.g. Endres, Bambach, & Floesser, 1971; Lima, Hale, & Myerson, 1991; Mortensen, Meyer, & Humphreys, 2006; Mysak & Hanley, 1959; Ramig, 1983, 1986; Ramig & Ringel, 1983). Investigations of age-related changes have often proceeded by comparing groups of young and old speakers (e.g. Duchin & Mysak, 1987; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008). On the one hand, such cross-sectional studies offer some important information not apparent in studies tracking individuals over time (cf. Braun & Friebis, 2009, for discussion). On the other hand, cross-sectional studies suffer from limitations: Just as cross-sectional demographic studies of many parts of the world would be consistent with the erroneous conclusion that people's ethnicity changes as they age, cross-sectional speech 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, Blondeau & Charity (2002), Trudgill (1988), Dubois (1992), Fowler (1986), and Mather (2011). 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, Maclagan, & Hay, 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 (Sancier & 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 at most three points in any participant's development.

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, Lancashire, Hirst, & Jokel, 2011) and the twenty-two letter writers (most of whom were famous authors) examined in Arnaud (1998). Analyses of vocal aging have similarly focused on figures of public life, including the Queen of England, former British Prime Minister Margaret Thatcher, the British-American journalist Alistair Cooke, and a number of other public figures (Braun & Friebis, 2009; Harrington, 2006; Reubold, Harrington, & Kleber, 2010). However, a limitation of these studies of famous individuals' linguistic usage 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. The difficulty inherent in collecting data in real time over long time spans – accompanied as it is by the researcher's own aging process – continues to stand in the way of systematic study of many variables of interest. In addition to the problem of obtaining longitudinal data, there is the problem of processing data in such a way as to enable systematic analysis. After all, freely-available archives of long-running soap operas, long-term TV and radio personalities, and so on, provide longitudinal speech samples of essentially unlimited size. Yet, the mere availability of such material does not ensure data that would readily allow linguistic analysis.

The current paper describes a project aimed at helping to remedy that scarcity of accessible pre-processed data. 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 (Apted, 1977, 1984, 1991, 1998; Apted & Almond, 2005).

A small number of previous studies have already made use of the Up series for sociolinguistic research. To our knowledge, the first study to do so (Sancier & Fowler, 1997) 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 transcriber agreement for vowels is known to vary substantially, based on studies comparing the coding decisions of multiple transcribers. For example, one study of a corpus of spontaneous speech (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005) reports agreement of 64% for [a], 67% for [ɑ], 55% for [ʊ], and 48% for [ʊ]. Of course, the task of transcribing all segments in a stretch of running speech (i.e. the task undertaken by the transcriber's in Pitt et al.'s study) differs from the binary classification task (e.g. [ʊ] vs. [ʌ]) in Sankoff (2004), and agreement on the latter task, had it been ascertained, may well have been substantially higher - even perfect. However, transcribers' coding decisions represent data about the transcribers' perception of a speech sample, rather than an unfiltered record of what the talker produced. Therefore, coded speech samples are maximally useful if accompanied by acoustic data and analysis. For example, rather than categorizing some tokens as "intermediate" between two extreme points, researchers may wish to study the variability of vowel quality, informing their judgment by an analysis of the variability along several acoustic dimensions. Two subsequent studies to make use of the Up series (Hansen, 2007; Poplack & Lealess, 2009) focused on two additional linguistic variables (h-dropping and theta-fronting) in 11 individuals. Neither study resulted in publicly available annotated speech corpora, but both underscore the usefulness of the films as linguistic data.

To address the need for more, and more fine-grained, data, 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. The purpose of the current paper is to describe the available data and to present an analysis of several acoustic parameters by way of illustrating some of the possibilities.

II MATERIALS

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; Apted & Almond, 2005). 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 the database, released in 2005, shows the participants at age 49.

III METHODS

We measured fundamental frequency (F0), the first four vowel formants, and speaking rate. Although the films largely consist of interviews and would thus seem to provide a very sizable speech sample, many parts of the films are difficult or impossible to analyze using the automatic alignment and formant extraction methods because of multiple talkers’ talking at once, or because of background noise: There are birds singing, trucks going by, or, at one point, a choir singing. Another constraint on the amount of analyzable data is that the films do not continually feature speech. The films often show the participants reacting to the interviewer's questions nonverbally or engaging in other activities besides talking. Wherever possible, we focused on utterances of at least 30 seconds or more of uninterrupted speech, 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, for example so as to analyze facial expressions, gaze direction, gestures, and any other visual information.

Once utterances were selected and coded for talker, age, and film, transcribers produced an orthographic transcript. For the formant analysis, the audio files were aligned with the transcripts at the phone level using the Penn Phonetics Lab Forced Aligner Toolkit (Yuan & Liberman, 2008). The Penn Forced Aligner is built on acoustic models of American English, while our sample is drawn from varieties of British English. While the aligning procedure yields vowel labels that differ from a narrow transcription of the data done by hand, such differences in labeling do not interfere with the analysis: Although the category labels may be different, for example in regards to the treatment of tense mid vowels as monophthongs vs. diphthongs, inspection of the results revealed that these differences did not affect the ability of the aligner to align the transcript with the signal: 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. 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, Hamakawa, Sakata, Hario, & Watanabe, 2007). The analysis reported here is based on the analysis frame occurring at the temporal midpoint of the vowel.

Speaking rate has been quantified in a variety of ways, including syllables per second (sometimes including pauses, sometimes excluding them), words per second, or, conversely, the average word or syllable duration. Following Ramig (1983), we quantified speaking rate as the average duration of a syllable. It should be noted that the syllable count was the based on the citation form of a word, rather than on the number of syllable peaks that were actually realized. In conversational speech, segments and whole syllables are frequently dropped (Johnson, 2004). The target syllable count therefore does not always equal the number of syllable peaks in the phonetic realization of a word. For example, the word “probably” is often realized with only two syllable peaks (occasionally rendered orthographically as “prolly”). This mismatch between citation syllable count and actual number of syllable peaks produced is problematic in some contexts, such as when the goal is to characterize phonetic syllable duration. However, when the goal is to estimate how fast a talker is talking, the target syllable count provides a useful measure of how quickly the talker is producing words in running speech. We report speaking rate as the average duration of syllables assuming the syllable count in the citation form. A range of alternative measures of speaking rate, such as words-per-second, can be straightforwardly extracted from the database.

IV CURRENT SIZE OF THE DATABASE

To date, we have analyzed 162 utterances (12,158 word tokens) produced by eight of the documentary participants, yielding 15,938 vowel tokens and 23,714 consonant tokens. For six of the participants (Andrew, Lynn, Nick, Paul, Sue, and Tony), we were able to find at least one utterance for every age represented in the films that met our inclusion criteria. Table I shows the utterance count, as well as the mean utterance length (in syllables), for each of the eight talkers at each age.

Table I: Number of length of utterances 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	Mean (and range of) syllable count per utterance	Number of utterances at each age				
			21	28	35	42	49
Andrew	15	71.9 (42-128)	1	2	4	6	2
John	14	163.6 (55-322)	4	n/a	6	n/a	4
Lynn	24	64.2 (41-116)	5	4	4	4	7
Nick	28	124.5 (43-388)	6	6	4	4	8
Paul	23	92.3 (38-193)	3	5	4	3	8
Sue	18	104.4 (62-212)	2	4	3	2	7
Symon	23	72.0 (32-119)	9	8	n/a	6	n/a
Tony	17	110.5 (35-254)	4	3	1	6	3

The database contains 12,158 word tokens, representing 1832 unique types, including 6278 tokens (1691 types) of open-class words. “Closed-class” words (or “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. Table II shows the word count for each of the eight talkers at each age. To date, the database contains a total of 39652 phones (23714 consonants and 15938 vowels). Table III shows the number of tokens for each phone, for each age.

Table II: Number of words produced by each talker at each age. Numbers in parentheses indicate the number of content words out of the total word count for each cell.

Talker	Total number of word tokens analyzed (total n = 12,158)	Number of word tokens at age				
		21	28	35	42	49
Andrew	791	55 (31)	139 (75)	218 (132)	306 (173)	73 (36)
John	1662	538 (295)	0	694 (344)	0	430 (234)
Lynn	1222	277 (137)	140 (77)	213 (107)	275 (152)	317 (156)
Nick	2543	363 (190)	803 (446)	460 (241)	461 (248)	456 (225)
Paul	1653	397 (177)	304 (164)	219 (107)	279 (150)	454 (225)
Sue	1467	151 (74)	276 (139)	277 (150)	278 (147)	485 (238)
Symon	1318	548 (279)	426 (213)	0	344 (176)	0
Tony	1502	452 (222)	192 (98)	93 (44)	489 (244)	276 (132)

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

Phone	Number of tokens produced at age ….				
	21	28	35	42	49
ɑ	116	135	129	132	134
æ	179	185	188	171	182
aɪ	339	217	200	279	268
aʊ	56	49	37	60	67
b	153	162	119	176	181
ɔ	140	71	87	101	102
ð	218	221	240	196	213
ʤ	51	42	34	50	43
d	388	338	294	308	357
ɛ	274	238	186	215	209
ɝ	159	186	170	170	168
eɪ	125	118	93	95	111
f	150	130	104	120	133
g	100	100	65	87	111
h	103	101	84	123	132
ɪ	453	427	396	382	437
i	315	285	280	355	378
j	110	98	64	81	66
k	241	210	178	231	224
l	317	281	284	277	296
m	286	227	218	285	261
n	611	532	502	537	552
ŋ	150	120	111	110	96
o	137	108	83	141	112
ɔɪ	14	8	7	3	2
p	156	137	144	135	136
r	326	276	268	278	287

ʃ	41	36	46	46	56
s	407	294	356	348	335
ʧ	38	45	36	35	38
t	701	606	590	675	689
ʊ	60	50	35	46	35
u	171	131	108	143	130
ʌ	983	884	879	898	921
v	156	165	159	170	159
w	217	198	184	186	239
z	199	174	194	216	201
ʒ	1	7	2	6	5
θ	98	64	62	56	52

V RESULTS

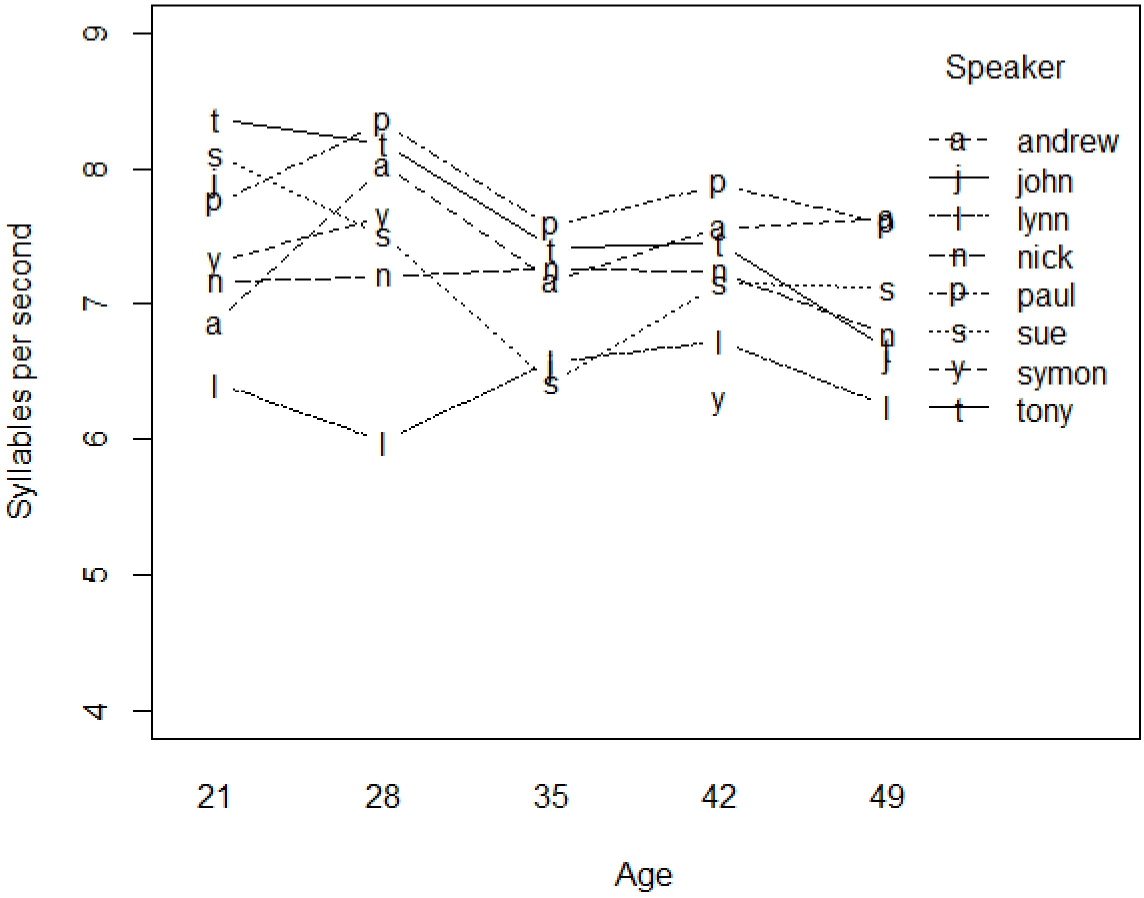
We now turn to two aspects of individual speaking characteristics that can be studied by means of the Up corpus: speaking rate and vowel space characteristics. We chose to focus on these measures because each of them has been shown to vary substantially within individuals or across groups and to act as a cue to the perception of age by listeners. In the case of vowel spaces, an additional motivation is the fact that acoustic analysis may reveal systematic subphonemic differences that are not necessarily detectable by transcribers.

Speaking rate

Speaking rate has been shown to be a significant cue to perceived age (Harnsberger et al., 2008). Consistent with this, some studies of acoustic correlates of aging tend to find a slowing of speech rate with increasing age, as one might perhaps expect, given the well-documented age-related decline in the speed with which adults perform a broad range of cognitive tasks (see e.g. Salthouse, 1991). Interestingly, this pattern is not always observed, with observations ranging from clear differences in the speaking rate of young vs. old adults (Duchin & Mysak, 1987) to more subtle ones (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001) down to an absence of any observable effect of age on speaking rate once other factors are controlled (Horton, Spieler, & Shriberg, 2010; Quené, 2008). One complication in studies of age-related change is that physiological age – a measure taking into account such measures as blood pressure, heart rate, vital air capacity, and general health – appears to be a more robust predictor of changes in vocal characteristics than is chronological age (Ramig, 1983, 1986; Ramig et al., 2001; Ramig & Ringel, 1983). For example, Ramig (1983) investigated speaking characteristics of speakers in two levels of physiological health ("good" and "poor") in three chronological age groups (age 25-35, 45-55, and 65-75 years) and found a clear pattern of reduced reading and speaking rates with advancing chronological age that was strongest in talkers in the "poor" physiological condition. Indeed, as Ramig (1983:224) points out "It is likely that if age-related changes in speaking and reading rates had been studied only in subjects in "good" physiological condition, no statistically significant effects would have been measured in this investigation." That observation must be taken as a cautionary note in interpreting studies taking into account talkers' chronological, but not physiological or biological age into account – particularly since many samples collected in university settings tends to be skewed towards "good" physiological condition – i.e. the part of the population in which speaking rate may not change perceptibly.

The average speaking rate (measured as the mean number of syllables per second) as a function of age, for each of the eight talkers, is shown in Figure 1.

Figure 1: Average number of syllables per second in the speech of eight talkers from ages 21 through 49.



As shown in Figure 1, speaking tempo, when measured as the average number of syllables per second, does not decrease steadily with increasing age for any of the speakers. However, it decreases between the ages 42 and 49 for each of the six talkers for whom data are available at those two ages, producing a decrease in the mean from 7.3 to 7.0 for the four male talkers, and a slight decrease from 6.9 to 6.8 for the two female talkers. This decrease is similar in magnitude to previously reported results (e.g. Ramig, 1983). The speaking rates observed here are overall higher than in that study, possibly in part because Ramig's sample was based on read speech. Another reason the syllable-per-second count appears high in the corpus, compared to the number of syllables per second typically observed in single-word production studies is that the measure plotted in Figure 1 takes into account function words, as well as content words. Since function words, as opposed to content words, are particularly susceptible to reduction and omission in conversational speech (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009), the resulting estimate of speaking rate appears high. Table IV shows the mean number of syllables per second when considering only content words. The pattern of results is similar to that observed when considering function words and content words together: Speaking rate decreases from age 42 to age 49 for all talkers for whom data are available, from 5.5 to 5.2 for the male talkers, and from 5.4 to 4.9 for the female talkers. Although the sample for each sex is small, it is worth noting that the decrease in speaking rate is present in the female talkers as well as the male ones when the syllable count is based on content words only. This pattern may reflect an increased reliance on function words.

Table IV: Mean syllable duration in content words produced by eight talkers at ages 21 through 49. Numbers in parentheses indicate standard deviations. Cells marked n/a indicate films in which talkers declined to participate.

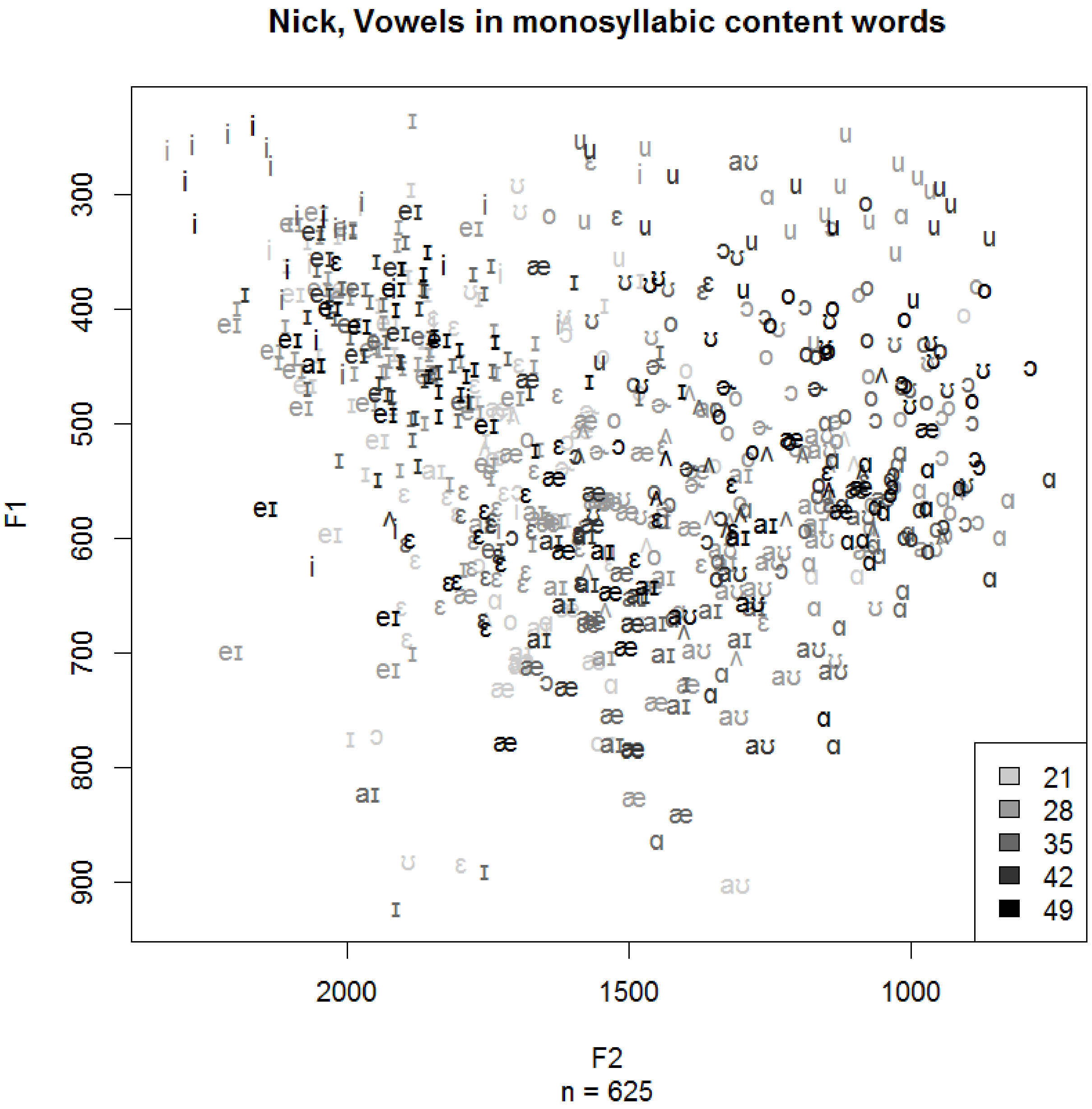
	21	28	35	42	49
Andrew	5.43 (2.44)	6.43 (4.32)	5.61 (2.01)	6.17 (3.06)	5.73 (2.37)
John	6.39 (3.07)	n/a	5.16 (2.67)	n/a	5.43 (3.79)
Lynn	4.98 (1.98)	4.98 (2.51)	4.92 (1.99)	5.45 (2.77)	4.72 (2.36)
Nick	5.01 (2.74)	5.2 (2.31)	5.29 (3.51)	5.48 (3.07)	4.86 (2.92)
Paul	5.15 (2.17)	6.04 (3.22)	5.65 (3.63)	5.73 (2.67)	5.47 (2.17)
Sue	5.72 (2.74)	5.19 (2.29)	4.88 (2.2)	5.26 (2.38)	5.07 (3.22)
Symon	5.53 (2.36)	5.68 (2.64)	n/a	5.02 (2.29)	n/a
Tony	5.28 (2.41)	5.48 (2.72)	4.89 (2.09)	5.1 (2.33)	4.94 (2.22)

Vowel spaces

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, Torretta, & Pisoni, 1996; Byrd & Saltzman, 2002; Clopper & Pierrehumbert, 2008; Clopper, Pisoni, & de Jong, 2005; Ferguson, 2007; Johnson, Flemming, & Wright, 1993; Lindblom, 1963, 1964). Age-related change in vowel spaces has 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; Linville, 2001). 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 (Linville, 2001).

Figure 2 shows the first two formant peaks in all tokens of vowels with primary stress in monosyllabic content words for one talker (“Nick”) from age 21 through 49. Increasing talker age is represented with increasingly dark plotting symbols. The overall pattern as it emerges in Figure 2 is that, with increasing talker age, the vowel space contracts, with vowel tokens getting increasingly likely to be produced near the center of the talker's vowel space.

Figure 2: Vowels in monosyllabic content words produced by one talker ("Nick") from ages 21 through 49.



VI DISCUSSION

The "Up" corpus provides a source of utterances from eight talkers recorded over a period of 28 years from young adulthood through old age. The small, but growing literature of longitudinal analyses of speech in adulthood has so far mostly focused on the speech of public figures and has offered little opportunity to study spontaneous speech of private individuals. The resource described here seeks to fill that gap.

Perhaps the most striking patterns that reveal themselves even in this sample analysis is that decreasing vowel dispersion and speaking rate, two features of speech associated with elderly talkers, are present during middle age adulthood, i.e. at a much earlier age than typically reported. In ongoing work, we are currently analyzing these patterns further.

Studies based on conversational speech data complement observations gleaned from formal interview situations or elicited word lists (Clopper et al., 2005; Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). 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, or to age-related physiological change.

VII CONCLUSION

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. In this article, we have demonstrated a few uses of the corpus. Studies making use of the corpus can easily extend the investigation to other acoustic variables known to play a role in age-related vocal change, such as fundamental frequency (Linville, 1988). It is hoped that the database will further stimulate the study of age-related and other individual variability.

Acknowledgments

This research was supported by the Committee on Research at UC Berkeley and by a Hellman Family Faculty Fund award to Susanne Gahl.

References

Apted, M. (Writer) (1977). 21 Up [Film]: ITV Granada Television.

Apted, M. (Writer) (1984). 28 Up [Film]: ITV Granada Television.

Apted, M. (Writer) (1991). 35 Up [Film]: ITV Granada Television.

Apted, M. (Writer) (1998). 42 Up [Film]: BBC.

Apted, M., & Almond, P. (Writer) (2005). 49 Up [Film]: ITV Granada Television.

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

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92-111.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Bradlow, A. R., Torretta, G., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Braun, A., & Friebis, S. (2009). Phonetic cues to speaker age: A longitudinal study. In G. Grewendorf & M. Rathert (Eds.), *Formal Linguistics and Law* (pp. 141-162). Berlin, New York: Mouton deGruyter.

Byrd, D., & Saltzman, E. (2002). Speech production. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 1072-1076). Cambridge: MIT Press.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America, 124*, 1682-1688.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America, 118*(3), 1661-1676.

Dubois, S. (1992). Extension particles, etc. *Language Variation and Change, 4*(2), 179-203.

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

Endres, W., Bambach, W., & Floesser, G. (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, S. H. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*, 1241-1255.

Fowler, J. (1986). The social stratification of (r) in New York City department stores, 24 years after Labov: New York University unpublished manuscript.

Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE Corpus. In J. C. Beal, K. P. Corrigan & H. L. Moisl (Eds.), *Creating and digitizing language corpora, Volume 2: Diachronic Databases* (pp. 82-104): Palgrave Macmillan.

Hansen, K. (2007). *H-Dropping as indicator of independent social variables*. Munich: GRIN Publishing GmbH.

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

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

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099-3111.

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

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated *Language, 69*, 505-528.

Le, X., Lancashire, I., Hirst, G., & Jokel, R. (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, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and aging, 6*(3), 416-425.

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

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

Linville, S. E. (1988). Intraspeaker variability in fundamental frequency stability: An age-related phenomenon? *Journal of the Acoustical Society of America, 83*(2), 741-745.

Linville, S. E. (2001). *Vocal Aging*: Singular.

Mather, P.-A. (2011). The social stratification of /r/ in New York City: Labov's Department Store Study Revisited. *Journal of English Linguistics*.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes, 21*(1-3), 238-290.

Mysak, E. D., & Hanley, T. D. (Writer) (1959). Vocal aging, *Geriatrics*.

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

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication, 45*, 89-95.

Poplack, S., & Lealess, A. V. (2009). Further insights from the "Up" series. Paper presented at *NWAV 38*. University of Ottawa.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America, 123*(2), 1104-1113.

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of communication disorders, 16*(3), 217-226.

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

Ramig, L. A., Gray, S., Baker, K., Corbin-Lewis, K., Buder, E., Luschei, E., et al. (2001). The aging voice: A review, treatment data and familial and genetic perspectives. *XXVth World Congress of the International Association of Logopedics and Phoniatrics, Aug 2001, Montreal, Canada; Portions of this research were presented at the aforementioned conference., 53*(5), 252-265.

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

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication, 52*(7-8), 638-651.

Salthouse, T. A. (1991). *Theoretical Perspectives on Cognitive Aging*. Hillsdale, NJ: Erlbaum.

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

Sankoff, G., Blondeau, H., & Charity, A. (2002). Individual roles in a real-time change: Montreal (r -> R) 1947-1995. In H. Van de Velde & R. van Hout (Eds.), *'r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/* (pp. 141-158). Brussels: ILVP.

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

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

Ueda, Y., Hamakawa, T., Sakata, T., Hario, S., & Watanabe, A. (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.

Yuan, J., & Liberman, M. (2008). *Speaker identification on the SCOTUS corpus.* Paper presented at the Proceedings of Acoustics 2008.