

# Between-speaker and within-speaker variation in speech tempo of American English

Ewa Jacewicz<sup>a)</sup> and Robert Allen Fox

*Department of Speech and Hearing Science, The Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210-1002*

Lai Wei

*Center for Biostatistics, The Ohio State University, 2012 Kenny Road, Columbus, Ohio 43210-1002*

(Received 8 December 2009; revised 11 June 2010; accepted 11 June 2010)

This study characterizes the speech tempo (articulation rate, excluding pauses) of two distinct varieties of American English taking into account both between-speaker and within-speaker variation. Each of 192 speakers from Wisconsin (the northern variety) and from North Carolina (the southern variety), men and women, ranging in age from children to old adults, read a set of sentences and produced a spontaneous unconstrained talk. Articulation rate in spontaneous speech was modeled using fixed-mixed effects analyses. The models explored the effects of the between-speaker factors dialect, age and gender and included each phrase and its length as a source of both between- and within-speaker variation. The major findings are: (1) Wisconsin speakers speak significantly faster and produce shorter phrases than North Carolina speakers; (2) speech tempo changes across the lifespan, being fastest for individuals in their 40s; (3) men speak faster than women and this effect is not related to the length of phrases they produce. Articulation rate in reading was slower than in speaking and the effects of gender and age also differed in reading and spontaneous speech. The effects of dialect in reading remained the same, showing again that Wisconsin speakers had faster articulation rates than did North Carolina speakers.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3459842]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Jt [DAB]

Pages: 839–850

## I. INTRODUCTION

The aim of this study is to characterize speech tempo of two distinct varieties of American English taking into account both within-speaker and between-speaker variation. Temporal changes in speech accompany any act of human communication being necessarily interconnected with the physical, social and psychological markings of speech. This omnipresent temporal variation can be further shaped by the demands of communicative situations, e.g., pertaining to more conversational or more formal speaking conditions. It has been long recognized that such conditional variations in speaking style affect not only general temporal characteristics of spoken utterances but introduce temporal changes to segmental properties. For example, vowel reduction or target undershoot is more likely to occur in faster, conversational speech (cf., Lindblom, 1963; Moon and Lindblom, 1994) whereas articulatory hyperarticulation is usually linked to “clear speech,” an intelligibility-enhancing speaking style which is manifested in slower speaking rate, more frequent pausing and general enhancement of segment-specific acoustic properties (cf., Picheny *et al.*, 1986; Fourakis, 1991; Perkell *et al.*, 2002; Krause and Braida, 2004; Smiljanić and Bradlow, 2008).

Apart from these and other intentional modifications to speech tempo, some individuals speak habitually faster (or

slower) than others (e.g., Tsao and Weismer, 1997; Tsao *et al.*, 2006). These individual differences reflect speaker-specific speed of articulatory movements and their unique use of prosody and pausing. Such within-speaker variation in speech tempo is additionally affected by a whole range of social and demographic variables such as pertaining to age, gender or geographic region of origin (Smith *et al.*, 1987; Amerman and Parnell, 1992; Byrd, 1994; Whiteside, 1996; Verhoeven *et al.*, 2004; Jacewicz *et al.*, 2009b). The latter factors are sources of the between-speaker variation (see Jacewicz *et al.*, 2009b, for further discussions and a more extensive literature review pertaining to the between-speaker variation).

Recent sociophonetic investigations involving larger corpora of American English and Dutch have brought to light systematic variations in speech tempo as a function of regional dialect, age and gender (Byrd, 1994; Verhoeven *et al.*, 2004; Quené, 2008; Jacewicz *et al.*, 2009b). In particular, variations in regional speech tempo were found which may be related to cultural differences (although the underlying causes are not well understood). In American English, there is a tendency for speakers from the north to speak faster than speakers from the south, which led to a stereotype of “slow-talking” Southerners and “fast-talking” Northerners (Niedzielski and Preston, 1999). One study examining speech rate in reading and frequency of pauses (Byrd, 1994) demonstrated that southern speakers may speak slower compared to northern speakers but attributed this difference, in part, to a more frequent use of pauses. Another study, Jacewicz *et al.*

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: jacewicz.1@osu.edu

(2009b) found that articulation rate (excluding pauses) of northern speech was significantly faster than that of southern speech, both in reading and speaking. Similarly, in Dutch, speakers from the Netherlands were shown to speak at a faster rate than those from Flanders (Verhoeven *et al.*, 2004; Quené, 2008). Regardless of the language background, younger adults in these American English and Dutch corpora tend to speak faster than older adults (see also Ramig, 1983; Smith *et al.*, 1987; Amerman and Parnell, 1992) and men tend to speak faster than women.

The present study aims to gain more insight into this type of variation in speech tempo of American English, asking further questions related to speaker dialect, age and gender. In particular, is there a systematic dialectal difference in speech tempo between the northern and southern regions? How does speech tempo change across the lifespan? Is male speech faster than female speech because of general gender-related differences or do the male and female speakers differ in their average phrase length, for example? Are differences in tempo as a function of the above factors manifested in both reading and spontaneous speech from the same individuals or is there a clear distinction between the articulation rate in reading and speaking?

This study is an extension of an earlier report by Jacewicz *et al.* (2009b) and focuses on the same regional dialects spoken in southeastern Wisconsin (Inland North, see Labov *et al.*, 2006) and western North Carolina (Inland South), representing northern and southern American English, respectively. These two regional varieties were selected because they exhibit two very distinct sets of northern and southern dialectal features in their respective phonological systems (Gordon, 2004; Thomas, 2004; Tillery and Bailey, 2004) as well as differences in their segmental durations (Jacewicz *et al.*, 2007; Jacewicz *et al.*, 2009a). In Jacewicz *et al.* (2009b), articulation rate in both read and informal speech was measured for men and women from two age groups of young and older adults (20–34 and 51–65 years, respectively) from these two regions. The focus was on between-speaker factors only, i.e., dialect, age and gender. Although the study found significant dialectal differences, the effects of age and gender were inconclusive. Given that within-speaker variation was not addressed and the age groups were narrowly defined, these results await further exploration.

The current study extends this investigation to address both between-speaker and within-speaker variation. Given the strong effect size of dialect in Jacewicz *et al.* (2009b), it is predicted that speech tempo of Wisconsin speakers will be faster than that of North Carolina speakers. To further explore the effects of age and gender, the study uses a larger corpus of 192 speakers and a wider age range, including children and adults up to their early 90s. In this paper, following Jacewicz *et al.* (2009b), speech tempo is defined as the articulation rate excluding pauses, measured in syllables per second, to express the number of output units per unit of time (Tsao *et al.*, 2006). By excluding the pause time, articulation rate conveys more closely the pace at which speech segments are actually produced and does not take into ac-

count speaker-specific ways of transmitting information, such as hesitations, pausing, emotional expressions, etc.

The examination of articulation rate is carried out in both a controlled condition (reading of sentences with a fixed number of syllables) and in spontaneous unconstrained talks. While the nature of the reading task allows us to examine the between-speaker variation in articulation rate without accounting for changes in phrase length which would be found in a larger read passage, the spontaneous talks provide a rich material to investigate more closely the within-speaker variation (see also a discussion of speaking versus reading rate in Jacewicz *et al.*, 2009b). The recent groundbreaking work of Quené (2008) opened a way to investigate this type of variation by means of multilevel/mixed-effects modeling (see also Goldstein, 2003; Quené and van den Bergh, 2004; Galwey, 2006; Bickel, 2007 for introduction of the mixed-effects models to applied research and tutorials). Mixed-effects models represent progress in the field of statistics and help to model the variation in spontaneous speech more adequately by overcoming the known limitations of the commonly used repeated-measures analysis of variance: homogeneity of variance, the sphericity assumption and treatment of missing data. Furthermore, the mixed-effects model allows for a variety of variance-covariance structures and not just compound symmetry as does repeated-measures ANOVA. For primary hypotheses testing, using the correct covariance structure is critical in producing unbiased estimates of standard errors.

That multilevel modeling is a more appropriate statistical tool for understanding speakers' variations in speech tempo in spontaneous speech has been shown convincingly in Quené (2008). This approach has more advantages than the repeated-measures ANOVA as it can account for the within-speaker variation. In this model, dialect, gender and age are included as between-speaker predictors (fixed effects) because these variables vary only between speakers and remain the same within speakers. However, individual speakers differ in the length of phrases they produce during their conversations (in the present sample, each speaker produced on average 59 phrases, ranging from 21 to 248). Presumably, this will introduce variability in the phrase length (a within-speaker variation), which necessitates including phrase length in the analysis and treating it as within-speaker factor. Yet, there is considerable variation in the manner that individuals convey a message, including frequency of pauses which determine the length of interpausal chunks of fluent speech used in the calculation of the articulation rate (a between-speaker variation). Therefore, phrase length has to be treated as both a fixed and a random effect. This variability is difficult to capture by repeated-measures ANOVA, leaving us with the mixed-effects model (multi-level modeling) as the most appropriate statistical tool.

As already discussed, speakers differ in how quickly or slowly they produce spontaneous speech. The nature of this variation depends on a number of factors including length of the utterance, discourse complexity, formality, affect, mood or communication style (e.g., Malécot *et al.*, 1972; Miller *et al.*, 1984; Duchin and Mysak, 1987; Walker, 1988; Ray and Zahn, 1990). Quené's (2008) results for 160 Dutch

TABLE I. Basic demographics of study participants. WI=Wisconsin, NC=North Carolina, m=male, f=female.

| Dialect region | Age group | Number and gender | Age range (years) | Mean age (s.d.) (years) |
|----------------|-----------|-------------------|-------------------|-------------------------|
| WI             | A0        | 10 m, 10 f        | 8–12              | 9.4 (0.8)               |
| NC             |           | 10 m, 10 f        |                   | 11.0 (1.5)              |
| WI             | A1        | 9 m, 9 f          | 20–34             | 23.7 (2.7)              |
| NC             |           | 9 m, 9 f          |                   | 26.8 (4.6)              |
| WI             | A2        | 8 m, 10 f         | 35–50             | 42.9 (4.3)              |
| NC             |           | 10 m, 10 f        |                   | 42.2 (4.9)              |
| WI             | A3        | 10 m, 10 f        | 51–65             | 57.2 (4.9)              |
| NC             |           | 10 m, 10 f        |                   | 58.3 (4.3)              |
| WI             | A4        | 9 m, 9 f          | 66–91             | 76.8 (6.2)              |
| NC             |           | 9 m, 9 f          |                   | 72.9 (7.4)              |

speakers showed that phrase length is a significant predictor of speech tempo so that longer phrases, containing more syllables compared to shorter ones, are generally spoken at a faster rate due to “anticipatory shortening” of the syllables. Taking into account the phrase length, more fine-grained analyses shed more light on the characteristics of Dutch speech. In particular, the speech of speakers from The Netherlands (the North) was shown to be faster than that of speakers from Flanders (the South). In addition, speakers in The Netherlands were found to produce shorter phrases than speakers in Flanders and older adults produced shorter phrases than younger adults. The present study, using a corpus of northern and southern American English speech, allows us to compare the regional variation in speech tempo of American English with that in Dutch. Since the study examines a wider range of speaker age including children and old adults, it broadens the experimental investigation typically carried out with young and middle-age adults. As in the Dutch corpus used in [Quené \(2008\)](#), the current work considers phrases and phrase length for individual speakers, which are built in the models developed for this study.

## II. METHOD

Two types of productions were elicited and analyzed in this study: a set of read sentences and a set of spontaneous unconstrained talks. In referring to the analysis of articulation rate of each different production type, the term *reading rate* will be used for the former and *speaking rate* for the latter.

### A. Articulation rate in reading (reading rate)

#### 1. Participants and data collection

Participants included 190 speakers who were born, raised and currently live in one of two distinct dialect regions in the United States: southeastern Wisconsin ( $N=94$ ) and western North Carolina ( $N=96$ ). The speakers fell into five age groups and included children aged 8–13 years (A0) and four groups of adults (A1, A2, A3, A4). The ages of the adult speakers ranged from 20 to 91 years. Each group spanned 15 years except for the oldest group that spanned 25 years. The age range, mean age and number of participants in each group are summarized in Table I. The recordings took place

in years 2006–2008. Each speaker represented the variety of American English typical of his or her dialect region as verified by the research staff. Speakers were recruited through posted advertisements on bulletin boards and in local newspapers, local radio announcements and through personal contacts in schools, nursing homes and local churches. They were paid for their participation.

Each speaker read a set of sentence pairs which were constrained contextually and prosodically in order to elicit variable emphasis of vowel production, which was a focus of a larger project. There were seven syllables in each sentence and the target vowels occurred in two contexts, [b\_ts] and [b\_dz]. The main sentence stress was systematically varied so that it could occur in five possible positions in the sentence. This was done for each vowel under study and for both consonantal contexts. A total of 240 sentences were read by adults in groups A1, A2 and A3. Children and the oldest adults (A4) read 120 sentences (half of the set) due to their increased fatigue with the task. Below are examples of sentence pairs used. A complete list of sentences can be found in Appendix I in [Jacewicz et al. \(2009b\)](#).

John said the BIG bets are low. No! John said the SMALL bets are low.

Sue KNOWS the small bites are deep. No! Sue THINKS the small bites are deep.

## 2. Data analysis

For the analysis of articulation rate, only the second sentence in the pair (underlined here) was chosen because it was produced more fluently than the first by most of the participants. The number of hesitations, “false starts” and pauses was greater in the production of the first sentence and it was felt that extensive editing of these utterances would compromise the measurement of the actual reading rate. Recording of these sentences was under computer control using a custom program written in Matlab. The sentence pairs appeared on a computer monitor in random order. The participant read the sentence pair speaking to a head-mounted microphone (Shure SM10A), placed at a 1-inch distance from the lips. For more details about recording procedure see [Jacewicz et al. \(2009b\)](#).

The reading time of each sentence was determined from the markings of the sentence onset and offset. The locations of sentence onsets and offsets were determined by hand from the waveform display using a waveform editing program (Adobe Audition). A reliability check was performed by a second researcher on all measurements using a custom Matlab program which displayed the onset and offset locations in the waveform. Average articulation rate in reading (per sentence) was measured in syllables per second, which was calculated by dividing the total number of syllables by sentence duration. In each sentence, there were a fixed number of seven syllables whose count was verified by the second researcher who performed the reliability check. The total number of included sentences was  $N=18\,229$  (11.12 h of speech). From the whole data set, 11 sentences were excluded due to wrong stress placement or misread words. There were also a small number of sentences which contained one or two pauses or hesitations (their proportion in the entire sample

was 0.03). These sentences were also included in the calculation of reading time after eliminating the pauses as it was felt this small number of slightly modified sentences would not introduce noise into the data. The number of sentences which were either excluded from the sample or edited was small because the original recording protocol required the experimenter to accept only those productions which met established criteria for stress placement, fluency, and absence of any obvious misarticulations. Multiple repetitions of a given utterance were allowed and only the best production (most fluent and with proper stress placement) was accepted for further analyses.

## B. Articulation rate in spontaneous talks (speaking rate)

### 1. Participants and data collection

The second type of recorded speech consisted of a short (up to 10 min) informal and unconstrained talk. The same speakers participated in the spontaneous talk task (and were recorded in the same testing session) except for young adult speakers in group A1. Because not all A1 speakers who read the sentence material also produced a talk, 6 new Wisconsin speakers and 12 new North Carolina speakers were added to the study. These speakers produced the talks only. A total of 38 speakers in A1 group produced informal talks (20 from Wisconsin and 18 from North Carolina, evenly divided by gender). In total, 192 speakers participated in the spontaneous talk task and their distribution across age groups A0, A2, A3 and A4 is listed in Table I.

The participants were recorded by the same experimenter as in the reading task, an adult female in Wisconsin and an adult female in North Carolina. Most speakers recounted stories from their lives or spoke about their families, friends, hobbies and their daily activities. They were instructed to speak for about 10 min at their typical tempo and mode, and that the topic of their talk was of their own choice. The talk was not intended to be an interview with the participant, which would necessitate a greater involvement of an interviewer. Rather, the goal was to elicit an uninterrupted discourse typical of that individual. In general, the talks produced by North Carolina speakers were longer than those produced by Wisconsin speakers.

### 2. Data analysis

For each speaker, two types of orthographic transcripts were prepared. In the first transcript, all words and sounds (such as hesitations or laughing) produced by the speaker were transcribed. The second transcript extracted interpausal phrases only, which were numbered consecutively. In this study, the phrase was defined as a string of words containing five or more syllables uttered without a pause. Next, the onset and offset of each phrase was measured using waveform editor (Adobe Audition) and the articulation rate was calculated in the same manner as for read sentences, i.e., by dividing the number of spoken syllables by the duration of each phrase. After listening to each phrase and marking its temporal onset and offset, the experimenter counted the number of syllables based on the spoken utterance (and not

on its orthographic notation). In total,  $N=11\,252$  phrases were analyzed from 192 speakers (7.22 h of speech). The number of included phrases per speaker ranged from 21 to 248 (mean=58.6, s.d.=25.2). As was done for the read sentences, a reliability check of both the phrase length and syllable count was performed by a second researcher using a dedicated custom Matlab program.

Articulation rate (per phrase, in syllables per second) was modeled using mixed-effects analyses (c.f., Pinheiro and Bates, 2000; Goldstein, 2003; Quené 2008; Quené and van den Bergh, 2008). The mixed-effects models were used to capture both within-speaker correlation and between-speaker variation. Four models were originally tested and the results of each optimal model were compared. Only the results of two optimal models are reported in this paper because the two remaining models did not yield significant insights into the variable of interest.

In the first model to be reported below, random effects were modeled at two hierarchical levels of speakers and phrases within speakers. None of the within speaker predictors (i.e., phrase length) were included. The fixed effects included the predictors of dialect, gender and age. In the second model, phrase length was included both in the fixed part and as a random effect nested within speakers. The two remaining optimal models which are not reported in this paper represented these same two models but with the addition of an interaction between dialect and age in the fixed part. This interaction was added as an exploration of the association of speaking rate with age according to dialect. Given that the participants varied in their educational background, the predictor of speaker's level of education (in number of years) was initially added to both models reported below, but in neither case was the model improved significantly.

Before modeling, the two levels of dialect were converted to a binary factor (North Carolina, Wisconsin) and speaker gender was also included as a binary factor (0 female, 1 male). Age was treated as a continuous variable (from 8 years to 91 years) and not as a factor based on the membership in the five age groups as in the read sentences task. Age was centralized to its mean. Furthermore, a quadratic relation between age and speech tempo was observed from the exploratory analysis of the data. Thus, *age quadratic* was used as another predictor to explore the best estimate of the predicted trend. For the second model, phrase length (in seconds of speaking time) was log-transformed and centralized to its mean log value. The sequential position of each phrase was not included in the models because in many cases, it was difficult to isolate a true "sequence" of consecutive phrases. Speakers produced hesitations, unexpected pauses and sequences of fillers between the phrases (such as "you know.... I really... I mean.... I was lost") which were edited out and this eliminated material introduced distortions to what might be considered a sequence of phrases. Given its potentially small effect found in Quené (2008), it was decided not to explore the sequential position in the current models.

For each model, the fixed part contains estimated regression coefficients ( $\beta$ ) and the random parts contain estimated amounts of variance between speakers ( $\sigma_u^2$ ) and between

phrases within speakers ( $\sigma_e^2$ ), both with standard errors. Mixed-effects models were performed using PROC MIXED procedure in SAS software (Version 9.1.3 of SAS © 2003 SAS institute Inc.), which uses a ridge-stabilized Newton-Raphson algorithm to optimize residual (REML) likelihood function to estimate the parameters. The  $t$  (or  $F$ ) statistics were used to test hypotheses of interest for the fixed effects, which account for the variance-covariance models we selected. For binary predictors, e.g., dialect and gender, and continuous predictors, e.g., age and phrase length, the significance of the estimated coefficients in the models was also tested by  $F$ -tests (the null hypothesis states that the estimated coefficient equals zero) and  $p$ -values are provided in the tables. Comparisons between the dialects (North Carolina and Wisconsin) and genders (female and male) were tested by the  $t$ -test (the null hypothesis states there is no difference between the testing groups). The mean of articulation rate was estimated for different dialect or gender group assuming average values for other predictors in the model.

### III. RESULTS

#### A. Reading rate

A univariate analysis of variance (ANOVA) was used to assess average articulation rate measured in syllables per second (syll/s) in read sentences, each of which contained seven syllables. The between-subject factors were dialect, age group and gender. The main effect of dialect was significant [ $F(1,170)=14.15$ ,  $p<0.001$ ] showing that Wisconsin speakers produced significantly more syllables per second (i.e., faster reading rate) than North Carolina speakers. The mean articulation rates were 3.31 syll/s (s.d.=0.399) for Wisconsin and 3.13 syll/s (s.d.=0.401) for North Carolina speakers.

The effect of gender was not significant although the general trend was that males spoke slightly faster than females. This trend was reversed in one age group in Wisconsin (A2) and two age groups in North Carolina (A1, A2), in which females spoke slightly faster than males. The latter effect was reflected in a significant three-way interaction between dialect, age group and gender [ $F(4,170)=3.25$ ,  $p=0.013$ ] although its effect size can be considered small as indicated by its low partial eta squared value ( $\eta^2=0.071$ ).

There was a significant effect of age group [ $F(4,170)=17.57$ ,  $p<0.001$ ]. Post hoc analyses were completed using Holm-Bonferroni-adjusted  $t$ -tests (Holm, 1979) for all pairwise comparisons between age groups. For each comparison between young adults (A1) and any of the four remaining age groups, the difference in the articulation rate was significant, showing that young adults read significantly faster (mean=3.58 syll/s, s.d.=0.44) than speakers from any other age group (the Holm-Bonferroni-adjusted  $p$ -values were: for A1-A0  $p<0.0001$ , for A1-A2  $p=0.0004$ , for A1-A3  $p=0.0002$ , for A1-A4  $p<0.0001$ ). The differences between the oldest adults (A4) and the remaining adult groups were also significant (for A2-A4  $p=0.0303$ , for A3-A4  $p=0.0321$ ), indicating that the oldest speakers read significantly slower (mean=3.0 syll/s, s.d.=0.35) compared to all other adults (but not children). None of the dif-

ferences between children (A0) and adults from groups A2, A3 and A4 were significant, which indicates that children's reading rate, although slower in general (mean=3.05 syll/s, s.d.=0.29), is comparable with that of adults with the exception of young adults (A1) as already discussed.

Overall, the results show significant dialectal differences in the reading rate: Wisconsin speakers read faster than North Carolina speakers. The effects of gender were not significant and the general trend for males speaking faster than females was inconsistent across the age groups. Differences as a function of speaker age were significant only for young adults who read faster than speakers from any other age group and for the oldest adults aged over 66 years who read significantly slower than the remaining adults. While the differences in the articulation rate as a function of dialect were apparent, the effects of age could be potentially linked to varying reading skills of the present speakers. In particular, most of the young adults were college students who were engaged extensively in reading on a daily basis, most likely more than speakers from any other age group. The oldest adults, on the other hand, could exhibit some reading problems related to their advanced age. In order to draw more conclusive results about the variation in articulation rate of these speakers, it is imperative to examine their tempo in spontaneous speech, unconstrained by the demands of the experimental task and potential reading problems. The modeling of spontaneous speaking rate was undertaken to overcome these limitations and provide a richer base on which we examine both the between-speaker and within-speaker variance.

#### B. Speaking rate

The first model given in Eq. (1) included the between-speaker factors dialect, gender and age (centralized) as fixed predictors. Both linear age and age quadratic were initially entered as predictors. However, because the linear term of age was not significant ( $p=0.78$ ) and a simpler model without the linear term of age yielded basically the same results, we decided to remove it and use the model that only included age quadratic. Random effects, i.e., speakers ( $u_{0j}$ ) and phrases within speakers ( $e_{ij}$ ), were modeled at two hierarchical levels, assuming the same phrase was not used by multiple speakers. This assumption was met by removing from analyses of speaking rate all phrases shorter than 5 syllables which are typically high frequency phrases such as "you know," "I mean," "I don't know." The talks were not based on a planned script nor an established passage. Each speaker talked about topics of interest to them and in an informal review of the transcripts we did not find any patterns of repeated phrases within speakers nor across speakers. There were no within-speaker predictors in the random part of the model. Random variances were assumed to be homogeneous.

$$y_{ij} = \text{dialect} \cdot \text{WI}[\gamma_{\text{dialect.WI } 00}] + \text{dialect} \cdot \text{NC}[\gamma_{\text{dialect.NC } 00}] \\ + \text{gender} \cdot \text{male}[\gamma_{\text{gender } 00}] + \text{age}^2[\gamma_{\text{age}^2 00}] \\ + (u_{0j} + e_{ij}). \quad (1)$$

TABLE II. Estimated parameters (with standard error) of multilevel modeling of articulation rate (syll/s).

| Model (1)           |          |                |                 |
|---------------------|----------|----------------|-----------------|
| Effects             | Estimate | Standard error | <i>p</i> -value |
| Fixed               |          |                |                 |
| dialect.Wisconsin   | 5.38     | 0.074          | <0.0001         |
| dialect. N.Carolina | 4.97     | 0.071          | <0.0001         |
| gender.Male         | 0.18     | 0.072          | 0.015           |
| age.Quadratic       | -0.00055 | 0.000070       | <0.0001         |
| Random              |          |                |                 |
| $\sigma_u^2$ $0_j$  | 0.22     | 0.025          | <0.0001         |
| $\sigma_e^2$ $ij$   | 1.08     | 0.015          | <0.0001         |

The results of model (1) are listed in Table II. First, there were significant dialectal differences showing that the speaking rate of Wisconsin speakers (estimated mean = 5.21 syll/s, s.e.=0.05) was significantly faster than of North Carolina speakers (estimated mean = 4.80 syll/s, s.e.=0.05) [ $t(189)=5.74$ ,  $p<0.0001$ ]. Second, males spoke significantly faster than females. Although the tempo difference between males and females was smaller than that between the dialects (the estimated mean for males was 5.09 syll/s (s.e.=0.05) and for females was 4.92 syll/s (s.e.=0.05) [ $t(189)=2.46$ ,  $p=0.015$ ]), the effect of gender turned out to be significant (which was not found for the reading rate). Third, the effect of age quadratic was significant. The predicted fixed effects of quadratic patterns for each dialect broken down by gender are shown in Fig. 1. As can be seen, the speaking rate increased as speaker age increased and achieved its peak value around the middle 40s, i.e., at 45 years-old. The amount of increase for each year increment of age is not constant but changes with age. The size of the increase decreases as the value of age approaches

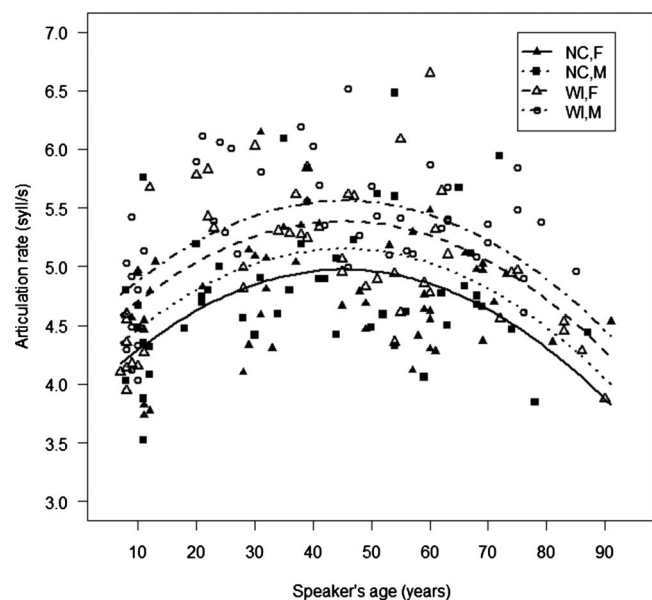


FIG. 1. The predicted fixed effects of quadratic patterns of age for each dialect broken down by gender. NC=North Carolina, WI=Wisconsin, M=males, F=females.

its peak value. After that point, speaking rate decreases with age and is slowest for the oldest speakers. The amount of decrease for each year increment of age is also not constant. Rather, the size of the decrement in speaking rate gets larger as the value of age gets farther from the peak value. From the random effects part, model (1) confirms that within-speaker variance is significant. The large size of the variance components suggests that more predictors may be added to the model.

However, before further modeling, we inquired into the relation between the average reading and speaking rates for the present speakers given that their reading rate was slower (3.20 syll/s, s.d.=0.39) compared to their speaking rate (mean = 4.96 syll/s, s.d.= 0.60 [ $t(169)=40.89$ ,  $p<0.0001$ ]). A linear regression analysis was used to model the relationship between the average speaking rate of each speaker onto his/her average reading rate. The results showed that reading rate was significant in modeling of speaking rate ( $p<0.0001$ ): Speakers who had a faster speaking rate had also a faster reading rate. As the reading rate increased 1 syll/s, the speaking rate increased 0.69 syll/s. This shows that there is a relationship between the articulation rate in speaking and in reading, which suggests the existence of the same underlying cause or a common motor control mechanism for speaker-specific rate of speech delivery.

Because of the nature of the reading task (each speaker producing the same set of utterances containing a fixed number of syllables, a completely crossed design), the within-speaker variance in reading could not be explored in this study. We focused therefore on further investigation of the within-speaker variance in speaking and developed a second model which builds on the results of model (1).

In the second model given in Eq. (2), phrase length (in seconds) was included as a fixed predictor in the fixed part. The model also contained the effects of phrase length in its random part nested within speakers. No additional predictors were contained in the model and the remaining predictors in the fixed part were as in model (1). Random variances were not assumed to be homogeneous.

$$\begin{aligned}
 y_{ij} = & \text{dialect} \cdot \text{WI}[\gamma_{\text{dialect.WI } 00}] + \text{dialect} \cdot \text{NC}[\gamma_{\text{dialect.NC } 00}] \\
 & + \text{gender} \cdot \text{male}[\gamma_{\text{gender } 00}] + \text{age}^2[\gamma_{\text{age}^2 \text{ } 00}] \\
 & + \text{phrase length}[\gamma_{\text{phrase } 00}] \\
 & + (u_{0j} + u_{\text{phrase } 0j} + e_{ij}). \tag{2}
 \end{aligned}$$

Table III shows the results of model (2). The results for the fixed part show that phrase length had a highly significant effect on speaking rate. Speakers produced shorter phrases with faster speaking rate as suggested by the negative coefficient of phrase length. The effects of dialect remained significant: Wisconsin speakers had a faster speaking rate (estimated mean=5.17 syll/s, s.e.=0.05) than North Carolina speakers (estimated mean=4.82 syll/s, s.e.=0.05) [ $t(189)=5.34$ ,  $p<0.0001$ ]. Significant effects of gender were also found. Males still produced significantly faster speaking rate than females (estimated means=5.07 (s.e.=0.05) and 4.92 syll/s (s.e.=0.05), respectively [ $t(189)$

TABLE III. Estimated parameters (with standard error) of multilevel modeling of articulation rate (syll/s). Model (2) includes the effects of phrase length (measured in seconds).

| Model (2)                              |          |                |                 |
|--|----------|----------------|-----------------|
| Effects                                | Estimate | Standard error | <i>p</i> -value |
| Fixed                                  |          |                |                 |
| dialect.Wisconsin                      | 5.36     | 0.068          | <0.0001         |
| dialect. N.Carolina                    | 5.00     | 0.065          | <0.0001         |
| gender.Male                            | 0.14     | 0.066          | 0.031           |
| age.Quadratic                          | -0.00054 | 0.000065       | <0.0001         |
| Phrase length                          | -0.48    | 0.039          | <0.0001         |
| Random                                 |          |                |                 |
| $\sigma_{u_{0j}}^2$                    | 0.19     | 0.022          | <0.0001         |
| $\sigma_{u_{length0j}}^2$              | 0.20     | 0.028          | <0.0001         |
| $\sigma_{u_{0j}}\sigma_{u_{length0j}}$ | -0.0077  | 0.018          | 0.67            |
| $\sigma_{e_{ij}}^2$                    | 1.0      | 0.014          | <0.0001         |

=2.17,  $p=0.031$ ], although this effect is reduced in comparison with model (1). Finally, the quadratic effect of age was significant and again in form as found in model (1): speaking rate increased with age achieving a peak at 45 years and decreased with age thereafter. As a whole, the results for the fixed part in Table III are similar to those in Table II, which supports the validity of the models.

The results for the random part revealed that speakers vary significantly in their average speaking rate according to the phrase length effect. The phrase length affected both the variance between speakers and within speakers and the within-speaker variation in average speaking rate was greater than the between-speaker variation. Figure 2 illustrates the between-speaker and within-speaker effects for phrase length ranging from 0.7 to 8.2 s (this includes more than 99% of phrases; extremely short and extremely long phrases with a small number of observations—less than 14—were removed). As can be inferred from Fig. 2, the variances between speak-

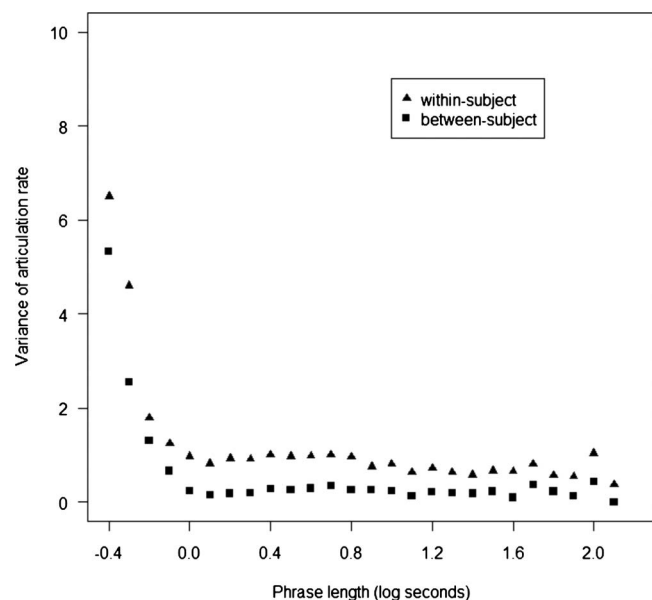


FIG. 2. Variance estimates for between-speaker and within-speaker variances broken down by phrase length.

ers decrease as phrase length increases, indicating that individual speakers tend to converge to the same speaking rate when they produce longer phrases. Furthermore, as phrase length increases, it tends to converge to the speaker’s average rate for that particular length with decreasing variation between phrases within speakers.

Model (2) brought to light the significance of the phrase length effect. Speakers have a faster speaking rate when their phrases are shorter. This finding is different from the results for Dutch in Quené (2008). In that study, Dutch speakers spoke faster when they produced longer phrases and this effect was attributed to “anticipatory shortening” of syllables in longer phrases. The present study shows the opposite in that these American English speakers had a faster speaking rate when their phrases were shorter. We will return to this in General Discussion. However, the finding that variances between speakers decrease as phrase length increases (with decreasing variation between speakers) is the same as in that study. Because the effect of gender (but not age) was smaller in model (2) ( $[F(1,189)=4.72, p=0.031]$ ), one possibility exists that the greater significance of gender in model (1) ( $[F(1,189)=6.07, p=0.015]$ ) arose due to the between-speaker effects on phrase length. It could be the case that males produce shorter phrases than females and these shorter phrases are produced at a faster speaking rate. Alternatively, male and female speakers may not differ in their average phrase length and the significant gender effects in models (1) and (2) may be due to some other gender-based differences in articulation rate such that males speak habitually faster than females. Model (3) was developed to address this and to model phrase length apart from speaking rate which was modeled in the two previous models.

However, before we present the results for model (3), we need to discuss measurement differences between the current study and Quené (2008). The measurement of speech tempo has taken basically two forms in the speech literature. First, it can be defined as the average rate at which speech units (such as syllables or words) occur in time, measured in syllables (or words) per second/minute (e.g., Stetson, 1951; Malécot *et al.*, 1972; Ramig, 1983; Trouvain and Grice, 1999; Tsao *et al.*, 2006; Verhoeven *et al.*, 2004). This is the older tradition—and, perhaps, the most commonly followed one—and is the approach taken in the current study. It provides a “global” (macro) characterization of the speech tempo used by a given speaker when producing an interpausal “phrase” (variously called a “run” or an “utterance” in other studies).

Alternatively, the inverse measure of syllables per second, average syllable duration (ASD) can be used (e.g., Miller *et al.*, 1984; Eefting, 1988; Crystal and House, 1990). ASD has often been used as a measure of speech tempo when acoustic characteristics of units in the phonological, prosodic or syntactic domains are examined (characteristics related to the structural complexity of the syllable, the effects of stress and rhythm, nature of the breath groups, etc.) which require syllable timing information in order to measure speech events at the micro level of detail. Quené (2008) opted to use ASD rather than syll/s as the measure of speech tempo. Mathematically, there is a perfect inverse relationship

between the two measures. [Campbell \(1988\)](#) notes that measurements of speech rates based on counts per unit time (or, alternatively, time per units) will be affected by the complexity of the units (e.g., the phonetic/phonological complexity of the syllable) but that this effect is reduced as the number of units increases and a more balanced distribution of unit types across the phrase (or run) occurs. We have reduced such effects here by eliminating analysis of all phrases less than 5 syllables in length.

Another difference between this study and [Quené \(2008\)](#) pertains to the measurement of phrase length: We have used the duration of the phrase (in seconds) while Quené used the number of syllables in the phrase. The choice of how to characterize the phrase length is, in large part, related to the choice of how to express speech rate. Our model includes a measure of phrase length expressed in terms of speaking time of the phrase in seconds as a factor in predicting speech tempo in syllables/second. Quené's model includes a measure of phrase length expressed in terms of the number of syllables in the phrase as a factor in predicting speech tempo in seconds/syllable.

Although we would expect the same basic result regardless of the approach taken (in terms of the significance and relative size of the predictive factors used here), it is an empirical question as to whether the same results would be obtained if we used ASD as the measure of speech tempo and number of syllables as the measure of phrase length. To answer this question, we converted syll/s to ASD and reran our model (2) using the number of syllables as the measure of phrase length. The pattern of results was practically identical (with significance levels for all effects remaining the same). Mean ASD for Wisconsin speakers (203 ms, s.e.=2.2) was significantly shorter than for North Carolina speakers (219 ms, s.e.=2.1), male speakers (208 ms, s.e.=2.2) had slightly shorter ASD than female speakers (214 ms, s.e.=2.2), and age quadratic was significant as was phrase length as a predictor. In our view, no matter which of the two approaches is taken in modeling overall speech tempo or phrase length as we do in model (3), the obtained results will be basically the same.

### C. Phrase length

The optimal model of phrase length is specified in Eq. (3). This model does not contain gender nor a quadratic age effect in its fixed part because gender and the quadratic term of age were not significant in its earlier versions. Consequently, model (3) contains only dialect and linear age (centralized) as predictors in the fixed part. Speakers ( $u_{0j}$ ) and phrases within speakers ( $e_{ij}$ ) were the two random effects. The phrase length (in seconds of speaking time) was log-transformed and centralized.

$$y_{ij} = \text{dialect} \cdot \text{WI}[\gamma_{\text{dialect.WI } 00}] + \text{dialect} \cdot \text{NC}[\gamma_{\text{dialect.NC } 00}] + \text{age}[\gamma_{\text{age}}] + (u_{0j} + e_{ij}). \quad (3)$$

Table IV lists the results of model (3). In the fixed part, we find that the coefficients for both Wisconsin and North Carolina dialects are significant. The results further show

TABLE IV. Estimated parameters (with standard error) of multilevel modeling of phrase length (in seconds, log-transformed and centralized).

| Model (3)           |          |                |                 |
|---------------------|----------|----------------|-----------------|
| Effects             | Estimate | Standard error | <i>p</i> -value |
| Fixed               |          |                |                 |
| dialect.Wisconsin   | -0.036   | 0.017          | 0.037           |
| dialect. N.Carolina | 0.033    | 0.017          | 0.045           |
| Age                 | -0.0011  | 0.00052        | 0.046           |
| Random              |          |                |                 |
| $\sigma_{u_{0j}}^2$ | 0.023    | 0.0028         | <0.0001         |
| $\sigma_{e_{ij}}^2$ | 0.21     | 0.0028         | <0.0001         |

that Wisconsin speakers produced shorter phrases (estimated mean=-0.036, s.e.=0.017) than North Carolina speakers (estimated mean=0.034, s.e.=0.017) [ $t(189)=-2.92$ ,  $p=0.004$ ]. The main effect of age was significant, showing a tendency for older speakers to produce shorter phrases than younger speakers. The log of phrase length decreases by 0.0011 for each 1-year increment of age. From the random effect part, model (3) confirms that both between-speaker and within-speaker variances are significant.

The results of the model of phrase length add to our understanding of the variation in the speaking rate as follows. The significantly faster speaking rate of Wisconsin speakers seems to be related to shorter phrases in their productions. By the same token, longer phrases produced by North Carolina speakers affect their speaking rate, which is significantly slower. No significant effects of gender on phrase length were found, indicating that phrases produced by male speakers are no shorter or longer than those produced by females. The differences between speaking rates of male and female speakers in models (1) and (2) thus seem to be related to some other gender-related differences and not to the length of their phrases. Finally, the effects of age on phrase length were significant, indicating that older speakers tend to produce shorter phrases than younger speakers. This finding can be interpreted in relation to the variation in speaking rate in the following way. Given the quadratic effect of age in models (1) and (2) (i.e., the speaking rate first increases with age achieving its peak value at 45 years-old, after which it decreases with age), the increase in the speaking rate may be related to the decrease in the phrase length up to the middle age. Since there was no significant quadratic affect of age on the phrase length, we can assume that speakers continue to produce increasingly shorter phrases as they get older and they also speak increasingly slower because of other age-related possible causes such as cognitive slowing or reduced speech motor control (e.g., [Bashore et al., 1998](#); [Zraick et al., 2006](#)). For example, older adults exhibit longer vowel and consonant durations and longer pauses between words which have a direct effect on the articulation rate (without pauses) and speech rate (including pauses) ([Benjamin, 1982](#); [Smith et al., 1987](#)). Thus, the relation between the faster speaking rates and shorter phrases may not exist for older speakers as it seems to be the case for younger



ones. However, we need to emphasize that the effect size for age in model (3) was very small, which does not allow us to draw firmer conclusions about age as a predictor of phrase length.

#### IV. GENERAL DISCUSSION

The present study measured the variation in articulation rate in American English as a function of regional dialect, age and gender using two different sets of speech materials: read sentences and spontaneous unconstrained talks. The major findings were that, in reading, the articulation rate of the northern speakers (from Wisconsin) was significantly faster than that of the southern speakers (from North Carolina) and young adults (aged 20–34 years) read significantly faster compared to any other age group. In spontaneous talks, the articulation rate of the northern speakers was again faster than of the southern speakers, male speakers were found to speak faster than female, and the speaking rate varied significantly across the lifespan: it first increased as speaker age increased, achieved its peak value around the middle 40s, and decreased with age thereafter being slowest for the oldest speakers. A closer examination of the phrase length effects revealed that speakers produced shorter phrases when they spoke faster, phrases produced by Wisconsin speakers were shorter compared to North Carolina speakers and older participants produced shorter phrases than younger. These results are discussed in greater detail below.

##### A. Variation in the reading rate

There is a substantial amount of research which addresses the variation in speech tempo in American English using read material, mostly structured utterances and longer passages of discourse (e.g., Crystal and House, 1982, 1988, 1990; Byrd, 1994; Tsao and Weismer, 1997; Tsao *et al.*, 2006). The choice of the reading materials varies across the studies and speech tempo is typically measured as a speaking rate (including pauses and hesitations) or articulation rate (excluding pauses and focusing on interpausal passages only). Although read text provides a valuable testing ground for examination of speech tempo since all speakers produce the same speech sample, it has serious drawbacks because speakers differ in their reading abilities and their reading styles. To reduce the interspeaker variation as a function of differences in reading including potential disfluencies, the study participants need to be selected more carefully and, in the published results, have usually been classified into the fast and slow talker groups (e.g., Crystal and House, 1982; Tsao and Weismer, 1997; Tsao *et al.*, 2006). Thus, the measured speaking rate (or articulation rate) should be in fact regarded as a reading rate which may differ from the actual speaking or articulation rate of the same individuals in tasks that do not require reading.

The present results give an indication that articulation rate in reading may be confounded by participants' reading abilities. First, the study found that the reading rate of the current young adults (20–34 years) was the fastest, which can be somewhat expected given that they were mostly college students engaged in extensive reading on a daily basis.

Second, the oldest adults' (over 66 years) rate was slower than that of the other adult groups perhaps because they were exhibiting some degree of age-related slowing in reading related to their poorer physiological conditions including slower cognitive processing times (e.g., Bashore *et al.*, 1989, 1998; Kail and Salthouse, 1994; Wingfield, 1996), general neuromuscular slowing (Ballard *et al.*, 2001) or some peripheral degeneration of the speech mechanism (Ramig, 1983; Smith *et al.*, 1987). Finally, a somewhat slower reading rate for children (although not significantly different from adults over 35 years old) might be expected as many may not have, as yet, reached adult levels of fluency in reading (especially the 8 year-olds).

Further confounding effects of reading ability on the articulation rate can be detected as a function of speaker gender. Based on the published reports for English, we may expect males to speak (or read) faster than females (e.g., Byrd, 1994; Whiteside, 1996). While this trend, although not significant, was also present in this study, the exceptions were found for adults aged 35–50 years in both dialects and for young adults (20–34 years) in North Carolina. In these groups, the females (and not males) read faster. These results are difficult to explain on the basis of gender-related differences in articulation rate.

However, it is unclear how the significant differences in the reading rate as a function of dialect are related to regional differences in reading. Rather, this effect comes from speech tempo differences between the northern and the southern speech (as also found earlier by Byrd, 1994) which are also manifested in reading.

##### B. Variation in the speaking rate

Consistent with the results for the reading rate, there was a robust difference between dialects in terms of the speaking rate in spontaneous talks: Wisconsin speakers had a significantly faster speaking rate than North Carolina speakers. The effects of age on speaking rate were somewhat different. While children and the oldest adults spoke slower than the remaining speakers, it was the adults in their 40s (and not young adults) that had the fastest speaking rate. Differences in speaking rate related to speaker gender were significant in models (1) and (2), indicating that in spontaneous talks (but not necessarily in reading), men speak faster than women. These results need to be further discussed in light of the systematic variation in the phrase length found in this study.

Considering the phrase length effects, the present results for American English are comparable to the results for spontaneous speech tempo in Dutch reported in Quené (2008). First, we find that phrase length differs between the speakers from Wisconsin and from North Carolina in the same way as it differs between the speakers from The Netherlands and from Flanders: speakers from the north (Wisconsin, The Netherlands) produce shorter phrases than those from the south (North Carolina, Flanders). Second, as in Dutch, phrase length did not vary with speaker's gender (and thus did not meet predictions stemming from our model (2)). The faster tempo for male speakers, both in American English and Dutch, seems to be related to gender differences in social

and linguistic behavior which may cause more systematic temporal variations in the production of speech segments. These variations can be related to differences in discourse structure, in social dominance and status, in phonetic consequences of participation in different stages of sound change and in a number of sociolinguistic variables (see [Cheshire, 2002](#), for a review). Third, older speakers in the present study were found to produce shorter phrases than younger speakers, which was also true for Dutch speakers. This general tendency exists despite differences in the age range of participants in the two corpora. Namely, Dutch speakers were all adults and, as school teachers, their ages could fall somewhere in between 25 and 65 years. The American English participants represented a wider range including children, young adults and old adults (from 66 years up to their early 90s). Thus, the age effects were expected to be present not only for phrase length but also as a between-speaker predictor of speaking rate. This was indeed the case and, unlike in the models for Dutch speakers, the effects of age for the American English speakers in models (1) and (2) persisted both when phrase length was ignored (model (1)) and when it was not ignored (model (2)).

An important discrepancy between the two studies pertains to the effects of phrase length. Although both studies found that speech tempo in a phrase strongly depends on the length of that phrase, the results went in an opposite direction. That is, longer phrases (containing more syllables) were spoken at a faster rate in Dutch but not in American English, where shorter phrases (containing less syllables) were found to be spoken faster. In an attempt to account for this discrepancy, we need to review the most relevant literature.

The effect of “anticipatory shortening” of syllables in longer phrases (as speakers anticipate more syllables to produce) than in shorter phrases is evoked to explain the results in the [Quené \(2008\)](#) study. This concept can be traced back to early studies such as [Schwartz \(1972\)](#) who examined three utterances: (1) /ipi/, (2) /ipi/ saw /ipi/, (3) /ipi/ saw /ipi/ with /ipi/ produced by four speakers and found that segmental duration (stop closure) decreased as the number of words in the utterance increased. He interpreted this result as a kind of anticipatory forward scanning which a speaker uses “to appraise the length of the utterance and uses this information to determine the amount of time he may devote to the articulation of individual sounds” (p. 666). Work exploring the relation between phrase length and phonological structure (e.g., [Jassem, 1952](#); [Lehiste 1972, 1977](#); [Jassem et al., 1984](#)) brought to light that a rhythm unit such as foot or anacrusis increases in duration linearly as a function of the number of syllables. This effect may be understood as an increase in speaking rate with longer phrase durations because at least some syllables in a rhythm unit are to be uttered more rapidly in order to maintain the rhythmic pattern of the utterance. Thus, the increased speaking rate is to be attributed to temporal reductions of syllables (and segments) within the rhythm unit and not an increase in speech tempo, per se. It is unclear whether and how these and similar findings, obtained in careful laboratory conditions and mostly using read speech, apply to unconstrained spontaneous utterances pro-

duced by speakers who vary in their use of syntactic structures, pauses, hesitations and emphasis, memory and vocabulary recall, etc.

For example, in spontaneous speech, speakers vary in the way they produce syntactically complex sentences and especially in the placement of pauses which produce interpausal phrases used in the calculation of the articulation rate. The pauses can occur in the middle of a sentence, in the middle of syntactic constituents or in other unpredictable locations ([Goldman-Eisler, 1968](#); [Grosjean et al., 1979](#)). The pause placement can further vary as a function of discourse organization and this variation can occur even in read speech ([Smith, 2004](#)). Given the considerable between- and within-speaker variation in producing interpausal phrases, it is unclear how much information is planned ahead in spontaneous speech when a phrase is produced. For example, the phrase from [Schwartz \(1972\)](#) “/ipi/ saw /ipi/ with /ipi/” can undergo variations such as “/ipi/, [you know], saw /ipi/ with /ipi/” or “/ipi/ saw [you know] /ipi/ with [well] /ipi/” where fillers like [you know] or [well] indicate hesitations and possible pausing. It is unclear which factors (cognitive or behavioral) determine the placement of pauses for different individuals who, in principle, should adhere to the same syntactic, phonological and prosodic structures of a given language.

The differences between [Quené \(2008\)](#) and the present study pertaining to the effects of phrase length could be related to differences in verbal planning of phrases produced by the two diverse participant pools. All participants in the Dutch corpus were school teachers who were accustomed to speaking in an instructional setting and this practice in the execution of syntactic and discourse complexity leads to a better linguistic skill ([Kess, 1992](#)). Their longer phrases might be produced at a faster rate due to their better command of spoken language in terms of articulatory planning, verbal monitoring and effective use of pauses whose appearance is reduced by practice ([Goldman-Eisler, 1968](#)). This, in turn, leads to a better control of syntactic complexity along with faster lexical retrieval ([Levelt, 1989](#), [Levelt et al., 1999](#)). To the contrary, the present American English speakers varied in their professional and educational background and in their experience with use of a spoken language. Although the level of education (in years) did not significantly affect their speaking rate as indicated by our initial modeling results, the fact that there were only 18 teachers in the North Carolina pool and 9 teachers in Wisconsin is in sharp contrast to the professional background of Dutch speakers. Furthermore, the nature of the present task (talking about their lives, families and recalling stories) is likely to produce a greater variation in the amount, the type and duration of hesitations and pauses, which may have an effect on their increased speaking rate in shorter phrases ([Kess, 1992](#)).

Another possibility is that language background may affect both speaking rate and utterance length, which may differ between Dutch and American English. In the present sample of spontaneous speech, the majority of American English phrases were rather short and the number of phrases containing from five to 12 syllables constituted 68% of all utterances. Interestingly, [Yuan et al. \(2006\)](#) reported a rapid rise of speaking rate of American English conversational

(telephone) speech for phrases containing from one to seven words (which could possibly contain up to 13–14 syllables) and no further increase for phrases containing from eight to about 30 words. Moreover, when corrected for the effects of phrase-final lengthening, the speaking rate of the 8-to-30-words phrases progressively decreased instead of remaining level. These findings correspond to the present results in that no increase in the speaking rate was found for longer phrases, yielding no support for the “anticipatory shortening” effect. A study by Nakatani *et al.* (1981) is often cited as providing evidence for such anticipatory shortening in English. However, this study has limited applicability to longer phrases and spontaneous speech because it examined temporal changes in English rhythm for reiterant speech phrases containing three, four and five syllables only.

The results of the present study clearly show that regional dialect can be a strong predictor of between-speaker variation in articulation rate. The present dialectal differences persisted when tempo measurements were corrected for the effects of phrase length in model (2). This suggests that dialect-specific features (rather than sources of within-speaker variation) account for cross-dialectal differences. These results are consistent with our previous research showing that North Carolina vowels are significantly longer than Wisconsin vowels, which is, in turn, related to distinct patterns of vowel dynamics in these two dialects in terms of vowel-inherent spectral change (Jacewicz *et al.*, 2007; Fox and Jacewicz, 2009). Thus, the longer vowels of North Carolina speakers are likely to produce longer syllables (i.e., slower articulation rate) and longer phrases compared to Wisconsin speakers. The present study of articulation rate adds to our better understanding of temporal differences among dialects by supplying evidence for systematic variation in both syllable and phrase duration.

## V. CONCLUSION

The present results for speech tempo in American English underscore the importance of an initial report by Miller *et al.* (1984) and subsequent findings by Quené (2008): There is a substantial variation in articulation rate across individual speakers and within utterances of each single speaker. While modeling this type of variation in spontaneous speech, the study found significant dialectal differences between northern and southern speech (including the length of phrases), significant changes in speech tempo across the lifespan and significant gender differences. Another finding was that articulation rate in spontaneous speech is faster than that in read speech and the two rates are correlated: Speakers who have a faster speaking rate have also a faster reading rate. In reading, however, the effects of between-speaker factors age and gender differed from those in spontaneous speech whereas the strong effects of dialect persisted. A comparison of the modeling results of spontaneous speech tempo in this study with those for Dutch (Quené, 2008) shows similarities with respect to northern and southern speech, age and gender effects. However, differences were found primarily for the effects of phrase length in that longer phrases (containing more syllables) were spoken at a faster rate in Dutch

but not in American English, where shorter phrases (containing less syllables) were found to be spoken faster. Further investigation is needed to address these discrepancies in terms of the lack of the apparent “anticipatory shortening” in American English as compared to Dutch and to explore several other predictors of spontaneous speech tempo.

## ACKNOWLEDGMENTS

This study was supported by the research grant No. R01 DC006871 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The authors would like to thank Joseph Salmons for his contributions to this research and the following for help with data collection, transcription and analysis: Mahnaz Ahmadi, Jason Fox, Sarah Hines, Anne Hoffmann, Yolanda Holt, Janaye Houghton, Katherine Lamoreau, Ting-fen Lin, Samantha Lyle, Caitlin O’Neill, Kristi Poole, Leigh Smitley, Dilara Tepeli, Lisa Wackler and Laura Winder.

- Amerman, J. D., and Parnell, M. M. (1992). “Speech timing strategies in elderly adults,” *J. Phonetics* **20**, 65–76.
- Ballard, K. J., Robin, D. A., Woodworth, G., and Zimba, L. D. (2001). “Age-related changes in motor control during articulator visuomotor tracking,” *J. Speech Lang. Hear. Res.* **44**, 763–777.
- Bashore, T. R., Osman, A., and Heffley, E. F. (1989). “Mental slowing in elderly persons: A cognitive psychophysical analysis,” *Psychol. Aging* **4**, 235–244.
- Bashore, T. R., Ridderinkhof, K. R., and van der Molen, M. W. (1998). “The decline of cognitive processing speech in old age,” *Curr. Dir. Psychol. Sci.* **6**, 163–169.
- Benjamin, B. (1982). “Phonological performance in gerontological speech,” *J. Psycholinguist. Res.* **11**, 159–167.
- Bickel, R. (2007). *Multilevel Analysis for Applied Research: It’s Just Regression!* (The Guilford, New York).
- Byrd, D. (1994). “Relations of sex and dialect to reduction,” *Speech Commun.* **15**, 39–54.
- Campbell, W. N. (1988). “Speech-rate variation and the prediction of duration,” In *Proceedings of the 12th International Conference on Computational Linguistics (COLING)*, Budapest, pp. 93–95.
- Cheshire, J. (2002). “Sex and gender in variationist research,” in *The Handbook of Language Variation and Change*, edited by J. K. Chambers, P. Trudgill, and N. Shilling-Estes, (Blackwell, Oxford), pp. 423–443.
- Crystal, T. H., and House, A. S. (1982). “Segmental durations in connected speech signals: Preliminary results,” *J. Acoust. Soc. Am.* **72**, 705–716.
- Crystal, T. H., and House, A. S. (1988). “Segmental durations in connected-speech signals: Current results,” *J. Acoust. Soc. Am.* **83**, 1553–1573.
- Crystal, T. H., and House, A. S. (1990). “Articulation rate and the duration of syllables and stress groups in connected speech,” *J. Acoust. Soc. Am.* **88**, 101–112.
- Duchin, S. W., and Mysak, E. D. (1987). “Disfluency and rate characteristics of young, middle-aged, and older males,” *J. Commun. Disord.* **20**, 245–257.
- Eefting, W. (1988). “Temporal variation in natural speech: Some explorations,” in *Proceedings of the Speech’88: Seventh FASE Symposium*, Edinburgh, pp. 503–507.
- Fourakis, M. (1991). “Tempo, stress, and vowel reduction in American English,” *J. Acoust. Soc. Am.* **90**, 1816–1827.
- Fox, R. A., and Jacewicz, E. (2009). “Cross-dialectal variation in formant dynamics of American English vowels,” *J. Acoust. Soc. Am.* **126**, 2603–2618.
- Galwey, N. W. (2006). *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance* (Wiley, Chichester).
- Goldman-Eisler, F. (1968). *Psycholinguistics. Experiments in Spontaneous Speech* (Academic, London).
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd ed. (Arnold, London).
- Gordon, M. J. (2004). “New York, Philadelphia, and other northern cities: Phonology,” in *A Handbook of Varieties of English*, edited by B. Kortmann and E. W. Schneider (Mouton de Gruyter, Berlin), pp. 282–299.

- Grosjean, F., Grosjean, L., and Lane, H. (1979). "The patterns of silence: Performance structures in sentence production," *Cognit Psychol.* **11**, 58–81.
- Holm, S. (1979). "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.* **6**, 65–70.
- Jacewicz, E., Fox, R. A., and Lyle, S. (2009a). "Variation in stop consonant voicing in two regional varieties of American English," *J. Int. Phonetic Assoc.* **39**, 313–334.
- Jacewicz, E., Fox, R. A., O'Neill, K., and Salmons, J. (2009b). "Articulation rate across dialect, age, and gender," *Lang. Var. Change* **21**, 233–256.
- Jacewicz, E., Fox, R. A., and Salmons, J. (2007). "Vowel duration in three American English dialects," *Am. Speech* **82**, 367–385.
- Jassem, W. (1952). *Intonation in Conversational English* (Polish Academy of Science, Warsaw).
- Jassem, W., Hill, D. R., and Witten, I. H. (1984). "Isochrony in English speech: Its statistical validity and linguistic relevance," in *Intonation, Accent and Rhythm. Studies in Discourse Phonology*, edited by D. Gibbon and H. Richter (de Gruyter, Berlin), pp. 203–225.
- Kail, R., and Salthouse, T. A. (1994). "Processing speech as a mental capacity," *Acta Psychol.* **86**, 199–225.
- Kess, J. F. (1992). *Psycholinguistics: Psychology, Linguistics and the Study of Natural Language* (John Benjamins, Amsterdam, Philadelphia).
- Krause, J. C., and Braid, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," *J. Acoust. Soc. Am.* **115**, 362–378.
- Labov, W., Ash, S., and Boberg, C. (2006). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin).
- Lehiste, I. (1972). "The timing of utterances and linguistic boundaries," *J. Acoust. Soc. Am.* **51**, 2018–2024.
- Lehiste, I. (1977). "Isochrony reconsidered," *J. Phonetics* **5**, 253–263.
- Levelt, W. J. (1989). *Speaking: From Intention to Articulation* (MIT, Cambridge, MA).
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). "A theory of lexical access in speech production," *Behav. Brain Sci.* **22**, 1–75.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Malécot, A., Johnston, R., and Kizziar, P. A. (1972). "Syllabic rate and utterance length in French," *Phonetica* **26**, 235–251.
- Miller, J. L., Grosjean, F., and Lomanto, C. (1984). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," *Phonetica* **41**, 215–225.
- Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 40–55.
- Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (1981). "Prosodic aspects of American English speech rhythm," *Phonetica* **38**, 84–105.
- Niedzielski, N., and Preston, D. (1999). *Folk Linguistics* (de Gruyter, Berlin).
- Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (2002). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," *J. Acoust. Soc. Am.* **112**, 1627–1641.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.* **29**, 434–446.
- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus, Statistics and Computing* (Springer, New York).
- Quené, H. (2008). "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *J. Acoust. Soc. Am.* **123**, 1104–1113.
- Quené, H., and van den Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Commun.* **43**, 103–121.
- Quené, H., and van den Bergh, H. (2008). "Examples of mixed-effects modeling with crossed random effects and with binomial data," *J. Mem. Lang.* **59**, 413–425.
- Ramig, L. A. (1983). "Effects of physiological aging on speaking and reading rates," *J. Commun. Dis.* **16**, 217–226.
- Ray, G. B., and Zahn, C. J. (1990). "Regional speech rates in the United States: A preliminary analysis," *Commun. Res. Rep.* **7**, 34–37.
- Schwartz, M. F. (1972). "Influence of utterance length upon bilabial closure duration for /p/," *J. Acoust. Soc. Am.* **51**, 666.
- Smiljanić, R., and Bradlow, A. R. (2008). "Temporal organization of English clear and conversational speech," *J. Acoust. Soc. Am.* **124**, 3171–3182.
- Smith, B. L., Wasowicz, J., and Preston, J. (1987). "Temporal characteristics of the speech of normal elderly adults," *J. Speech Hear. Res.* **30**, 522–529.
- Smith, C. (2004). "Topic transitions and durational prosody in reading aloud: Production and modeling," *Speech Commun.* **42**, 247–270.
- Stetson, R. (1951). *Motor Phonetics: A Study of Speech Movements in Action*, 2nd ed. (North-Holland, Amsterdam).
- Thomas, E. (2004). "Rural southern white accents," in *A Handbook of Varieties of English*, edited by B. Kortmann and E. W. Schneider (Mouton de Gruyter, Berlin), pp. 300–324.
- Tillery, J., and Bailey, G. (2004). "The urban South: Phonology," in *A Handbook of Varieties of English*, edited by B. Kortmann and E. W. Schneider (Mouton de Gruyter, Berlin), pp. 325–337.
- Trouvain, J., and Grice, M. (1999). "The effect of tempo on prosodic structure," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, pp. 1067–1070.
- Tsao, Y.-C., and Weismer, G. (1997). "Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component," *J. Speech Lang. Hear. Res.* **40**, 858–866.
- Tsao, Y.-C., Weismer, G., and Kamran, I. (2006). "Interspeaker variation in habitual speaking rate: Additional evidence," *J. Speech Lang. Hear. Res.* **49**, 1156–1164.
- Verhoeven, J., De Pauw, G., and Kloots, H. (2004). "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands," *Lang Speech* **47**, 297–308.
- Walker, V. G. (1988). "Durational characteristics of young adults during speaking and reading tasks," *Folia Phoniatr.* **40**, 12–20.
- Whiteside, S. P. (1996). "Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences," *J. Int. Phonetic Assoc.* **26**, 23–40.
- Wingfield, A. (1996). "Cognitive factors in auditory performance: Context, speech of processing and constraints of memory," *J. Am. Acad. Audiol.* **7**, 175–182.
- Yuan, J., Liberman, M., and Cieri, C. (2006). "Towards an integrated understanding of speaking rate in conversation," in *Proceedings of the Interspeech-2006*, Pittsburgh, PA, pp. 541–544.
- Zraick, R. I., Gregg, B. A., and Whitehouse, E. L. (2006). "Speech and voice characteristics of geriatric speakers: A review of the literature and a call for research and training," *J. Med. Speech-Lang. Pathol.* **14**, 133–142.