Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo

Hugo Quené[a)]
Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

(Received 11 October 2006; accepted 7 November 2007)

Speech tempo (articulation rate) varies both between and within speakers. The present study investigates several factors affecting tempo in a corpus of spoken Dutch, consisting of interviews with 160 high-school teachers. Speech tempo was observed for each phrase separately, and analyzed by means of multilevel modeling of the speaker's sex, age, country, and dialect region (between speakers) and length, sequential position of phrase, and autocorrelated tempo (within speakers). Results show that speech tempo in this corpus depends mainly on phrase length, due to anticipatory shortening, and on the speaker's country, with different speaking styles in The Netherlands (faster, less varied) and in Flanders (slower, more varied). Additional analyses showed that phrase length itself is shorter in The Netherlands than in Flanders, and decreases with speaker's age. Older speakers tend to vary their phrase length more (within speakers), perhaps due to their accumulated verbal proficiency. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821762]

PACS number(s): 43.70.Fq, 43.70.Jt, 43.70.Bk, 43.71.Bp [BHS] Pages: 1104–1113

I. INTRODUCTION

Speech tempo may be defined as the rate at which phonetic events occur over time. It is often expressed in phonemes or syllables per second (e.g., Stetson, 1951), or inverted, as seconds per syllable [average syllable duration (ASD), e.g., Goldman-Eisler (1968), Crystal and House (1990)]. Humans do not produce speech at a constant tempo (e.g., Miller *et al.*, 1984b). Speech tempo is reportedly affected by various factors, some varying within speakers, e.g., length of phrase, and others varying between speakers, e.g., sex. The present study aims to investigate several of these factors affecting tempo, by means of a corpus of spontaneous Dutch produced by 160 speakers from The Netherlands and Flanders (Belgium). A secondary purpose is to demonstrate how multilevel models can be used to capture effects varying at multiple hierarchical levels: that of speakers, and of phrases within speakers.

Speech tempo is defined here as the articulation rate, excluding pause time. In the present analysis, speech tempo is computed by taking the duration of each interpausal phrase in the speech corpus, and dividing that duration by the number of canonical or intended syllables in that chunk (Koreman, 2006), yielding ASD for that phrase.

A previous study of speech tempo in Dutch (Verhoeven *et al.*, 2004), using the same Dutch corpus material as in the present study, found significant effects of four predictors varying between speakers, viz., speakers' age, sex, country, and region. Speakers' average tempo was reported to be faster for younger speakers (ages 21–40) than for older speakers (ages 45–59), and faster for men than for women (Verhoeven *et al.*, 2004). For speakers of Dutch from The Netherlands, four dialect regions were investigated (Verhoeven *et al.*, 2004). The Randstad region (Zuid-Holland) is considered the linguistic center of The Netherlands. The Mid-region (Utrecht, Gelderland) is a transition zone. The North (Groningen, Drenthe) and South (Dutch Limburg) regions have distinct regional dialects, although the "western" variety of Standard Dutch is widely used. For speakers of Dutch from Flanders (Vlaanderen), four dialect regions were also investigated. The Brabant region is considered the linguistic center of Flanders. East Flanders is regarded as a transition zone, whereas Belgian Limburg and West Flanders are regarded as more peripheral. Speakers' average tempo in the standard variety of the language was found to be related to their distance from the "linguistic center." For speakers from The Netherlands, average tempo was fastest in the Randstad region, intermediate in the Mid-region, and slowest in the North and South regions (Verhoeven *et al.*, 2004). For speakers from Flanders, regional differences were not significant. On average, speech tempo was considerably slower in Flanders than in The Netherlands.

This previous analysis of the Dutch corpus material, however, was limited to *between-speaker* variation in tempo. In practice, however, within-speaker changes in tempo are probably more relevant for speech communication. The present study therefore extends previous models of Dutch speech tempo, by including several *within-speaker* predictors in the model.

Longer phrases, containing more syllables, tend to be spoken at a faster rate, with shorter average syllable durations; this is known as "anticipatory shortening" (Nooteboom, 1972; Lindblom and Rapp, 1973; De Rooij, 1979; Nakatani *et al.*, 1981). A plausible model of speech tempo should therefore include phrase length as a predictor, to capture the considerable within-speaker variation (Miller *et al.*, 1984b). If speakers from different regions would differ in their average phrase length, for example, then such differences would yield artifactual differences in speech tempo,

[a)]Electronic mail: hugo.quene@let.uu.nl.

through the mechanism of anticipatory shortening. Hence, phrase length has to be accounted for when studying speech tempo in corpus materials.

The speech corpus under analysis consists of interviews with high-school teachers of Dutch language and literature. Each interview lasted about 15 min. During the interview, speakers may gradually speed up (e.g., due to arousal) or slow down (e.g., due to fatigue). Hence, the sequential position of a phrase constitutes a second within-speaker predictor of the speech tempo within that phrase.

There may be cross correlations in tempo among adjacent phrases within each interview, e.g., because speakers may tend to alternate slow and fast responses. In order to investigate such possible autocorrelations the average speech tempo of a previous phrase was also included as a predictor. This was done for lags of one to five phrases, yielding five autocorrelation predictors.

Unfortunately, the present corpus material does not allow investigations of two other within-speaker factors that may well be relevant, viz., emphasis and emotional involvement. Many textbooks in phonetics state that speakers vary their speaking rate, in anticipation of the time listeners will need to process their words. Hence, important or unpredictable portions are typically spoken at a relatively slower rate (Zwaardemaker and Eijkman, 1928; Nooteboom and Eefting, 1994), because "we speak…in order to be understood" (Jakobson and Waugh, 1979, p. 95). Both emphasis and emotional involvement are very difficult and costly to annotate, however, and such annotations are therefore not available for any large-scale corpus such as used in this study.

The primary goal of this study is to adequately model spontaneous speech tempo, and speakers' variation in tempo, using a large corpus of spontaneous speech. These variations are relevant in speech communication, because listeners use the input speech tempo as a scaling factor for other phonetic distinctions, such as the distinction between /ba/ vs /wa/ (e.g., Miller *et al.*, 1984a), voicing of consonants (e.g., Miller and Grosjean, 1981; Volaitis and Miller, 1992), and quantity of vowels (e.g., Traunmüller and Krull, 2003). If there is a lot of within-speaker tempo variation, then this might make a listener's task more difficult, because the listener has to readjust and rescale the phonetic distinctions more frequently. The secondary goal of this study is to illustrate how multilevel modeling may help us achieve the first goal, by simultaneously modeling both tempo and tempo *variations*, at multiple hierarchical levels.

II. METHOD

The Corpus of Spoken Dutch (Oostdijk, 2000) was used to investigate which factors contribute to (variation in) speaking rate. For this purpose, we concentrated on the subcorpus containing interviews with $N=160$ high-school teachers of Dutch in Flanders (the Dutch-speaking part of Belgium) and in The Netherlands. Interviewed speakers ("interviewees") were stratified by dialect region (four regions within The Netherlands, and four within Belgium), sex, and age group (below 40 versus over 45 years of age), with $n=5$ speakers in each cell. All speakers are assumed to speak a variety of Standard Dutch. The 80 interviews within each country were conducted by the same interviewer (Netherlands: male, age 26; Belgium: female, age 23). (Using a single interviewer across all interviews would have resulted in varying degrees of cross-cultural difference between conversation partners.) Similar topics were discussed across all interviews. Hence, language variety, conversation partner, and conversation topics were largely eliminated as confounding factors, and the speech samples were highly comparable among speakers. Nevertheless, any differences in results between The Netherlands and Belgium may also be ascribed to different interviewers (conversation partners) being used in the two countries. Note, however, that the sex difference among conversation partners is balanced in both countries, and that both interviewers have approximately the same age. Hence any differences observed in this corpus study are most likely due to cultural differences between the two countries, and effects of the two interviewers' properties are further ignored in this study. For more details about dialect regions, speaker selection, and recording procedure, see Van Hout *et al.* (1999), Verhoeven *et al.* (2004), and Adank *et al.* (2007).

For each interview, the orthographic transcripts were extracted from the annotations provided with the corpus, broken down by interpause chunks. The speaking time of each chunk or phrase was determined from the time marks in the transcript. The number of orthographic syllables in each phrase was determined by dictionary look-up of the orthographic words (with manual correction where necessary). Speech tempo is expressed here as ASD (Goldman-Eisler, 1968; Crystal and House, 1990). Interpause chunks or phrases, as determined by the original transcribers, constitute the smallest units of observation. Hence, the average syllable durations per phrase correspond to so-called articulation rates, from which pause time is excluded. Most of the short phrases (of one or two orthographic syllables) consisted of hesitation sounds, filled pauses, backchannel sounds, etc. These were excluded from the data set [after Crystal and House (1990)]. The proportion of excluded phrases per speaker ranged from 0.01 to 0.27 (median 0.08, interquartile range 0.08). The number of included phrases per speaker ranged from 139 to 629 (median 328, interquartile range 97), with corresponding speech periods ranging from 4 to 31 min of speech per speaker (median 13.6 min, interquartile range 4.7). In total, $N=52\,975$ phrases from 160 speakers were included in the analyses (38.5 h of speech).

Average syllable durations (per phrase) were modeled by means of multilevel analysis (Goldstein, 1995; Cnaan *et al.*, 1997; Kreft and De Leeuw, 1998; Snijders and Bosker, 1999; Pinheiro and Bates, 2000; Luke, 2004), with speakers and phrases within speakers as two nested random factors. This type of analysis has several important advantages over more conventional techniques such as repeated measures analysis of variance (ANOVA), or ordinary least-squares linear regression [see Quené and Van den Bergh (2004) for a longer review and tutorial]. First, it allows for multiple, nested random effects, such as speakers and phrases within speakers. Second, multilevel modeling does not require homogeneity of variance, or sphericity. Between-speaker variance s_u^2 and within-speaker variance s_e^2 are modeled explicitly, instead of assumed to be homogeneous. This property allows us to investigate whether these between-speaker and within-speaker variances are affected by any of the predictors under study. Third, multilevel modeling allows for incomplete designs, and for varying numbers of observations per cell. This property removes the necessity to aggregate multiple lower-level observation units (phrases) into a single value at the higher level (speakers), as was done by Verhoeven *et al.* (2004). The present multilevel analysis assumes that phrases are nested under speakers, i.e., that phrases are not repeated between or within speakers. Although a few phrases were repeated (e.g., *ja precies* "yes exactly"), the number of these repeated phrases in the corpus was sufficiently low (only 147 out of 52 975) to warrant this assumption.

Before multilevel modeling, the eight levels of the region factor were converted to eight binary (dummy) factors. Sex was also included as a binary dummy factor (0 female, 1 male). Age was not included as a discrete factor discriminating two speaker groups [like Verhoeven *et al.* (2004)], but as a linear predictor, centralized to the mean age of 43 before modeling (Snijders and Bosker, 1999; Kreft and De Leeuw, 1998; Luke, 2004). Phrase length was log-transformed and centralized to its mean log value; the sequential position of each phrase was also centralized.

In any type of statistical modeling, the aim is to obtain an optimal model that contains the lowest number of predictors but explains the highest amount of variance of the dependent variable. In multilevel modeling, this is complicated somewhat because the optimal model can be different for the fixed part, and for the random parts at each level. Therefore one may find different predictors in the various parts of the models. For each model reported in the following, the fixed part contains estimated regression coefficients (β), and the random parts contain estimated amounts of variance between speakers (σ_u^2) and between phrases within speakers (σ_e^2), with standard errors for the former estimates, and confidence intervals for the latter estimates. The optimal model was selected by means of comparisons among candidate models with and without relevant effects in their random parts, using likelihood ratio tests (Pinheiro and Bates, 2000). Of the multiple candidate models for a particular set of predictors, only the optimal one will be reported in the following, for the sake of clarity.

The resulting estimates can be used for testing hypotheses. For binary predictors, e.g., sex, and for continuous predictors, e.g., age, the H_0 states that the estimated coefficient equals zero. Such hypotheses may be tested by using the Wald criterion that the estimated coefficient is significant if the coefficient exceeds 1.96 times its standard error [at $\alpha=0.05$; Hox (1995), Snijders and Bosker (1999)]. Other factors are modeled with multiple binary dummy predictors, e.g., region consisting of one binary predictor for each of the four regions. The main effect of the region is investigated by means of a planned contrast among the estimated coefficients for these binary predictors, with H_0 stating that this contrast equals zero. The amount of variance associated with such a contrast is then tested by means of a χ^2 distribution (Winer, 1971; Goldstein, 1995). Random estimates are evaluated by means of Markov chain Monte Carlo sampling; this is a Bayesian technique yielding estimates of the posterior distributions of all parameters, from which the 95% confidence of the random estimate may be extracted (Browne, 2005).

III. RESULTS

A. Speech tempo

The first model fitted was deliberately similar to the between-speaker ANOVA reported by Verhoeven *et al.* (2004) for the same corpus. This preliminary model given in Eq. (1) included only the between-speaker factors country, region, sex, and age (centralized) as fixed predictors. Random effects (in parentheses) were modeled at two hierarchical levels, viz., that of speakers (u_{0j}) and that of phrases within speakers (e_{ij}). In other words, articulation rates were not aggregated at the speaker level in the present study,

$$\begin{aligned} y_{ij} = {} & \text{reg.NR}[\gamma_{\text{NR }00}] + \text{reg.NM}[\gamma_{\text{NM }00}] \\ & + \text{reg.NN}[\gamma_{\text{NN }00}] + \text{reg.NS}[\gamma_{\text{NS }00}] \\ & + \text{reg.FB}[\gamma_{\text{FB }00}] + \text{reg.FE}[\gamma_{\text{FE }00}] \\ & + \text{reg.FL}[\gamma_{\text{FL }00}] + \text{reg.FW}[\gamma_{\text{FW }00}] \\ & + \text{sex.Male}[\gamma_{\text{sex }00}] + \text{age}[\gamma_{\text{age }00}] + (u_{0j} + e_{ij}). \end{aligned} \qquad (1)$$

None of the within-speaker predictors were included, and random variances were assumed to be homogeneous, i.e., the predictors were not included in the random part of the model. The results of this preliminary model are listed in the left-hand part of Table I.

Results for this preliminary model (1) confirm previous analyses of speakers' average tempo in this corpus (Verhoeven *et al.*, 2004). First, comparisons of the regional means show that speakers from Flanders produced significantly longer syllables (i.e., slower tempo) than did those from The Netherlands [$\chi^2(1)=201.3$, $p<0.001$]. Second, within The Netherlands, speakers from the Randstad region (the linguistic center of The Netherlands) produced significantly shorter syllables (i.e., faster tempo) than did those from the other regions [$\chi^2(3)=12.2$, $p=0.007$]. Within Flanders, the corresponding contrasts between Brabant and the other regions are marginally significant [$\chi^2(3)=7.5$, $p=0.058$]. Third, male speakers produced significantly shorter syllables (i.e., faster tempo) than female speakers ($Z=-3.70$, $p<0.001$). Fourth, older speakers produced significantly longer syllables (slower tempo) than younger speakers ($Z=4.24$, $p<0.001$). For each additional year of age, ASD increases by 0.75 ms. With a grand mean ASD of 236 ms, the tempo difference between speakers aged 25 and 65 is $(40\times 0.75)/236$ or about 13%. This age effect is well above the just noticeable difference (JND) for speech tempo of about 5% (Quené, 2007). The between-speaker variance is comparable to that reported in previous studies. Speakers' average syllable duration ranged from 177 to 333 ms (mean 238 ms, s.d. 36 ms). Tsao *et al.* (2006) compared the 15 slowest and the 15 fastestspeakers, in a sample of 100 speakers, and reported vowel durations of the slowest speakers that were

TABLE I. Estimated parameters of multilevel modeling of syllable durations (in ms). Estimates of fixed parameters are given with standard error (in parentheses); estimates of random parameters are given with 95% confidence intervals obtained from Markov chain Monte Carlo sampling (in parentheses).

	Model (1)		Model (2)	
Fixed				
reg.N.Randstad	204.9	(5.5)	193.1	(3.5)
reg.N.Mid	224.1	(5.5)	201.1	(3.6)
reg.N.North	225.2	(5.5)	193.5	(3.5)
reg.N.South	227.1	(5.5)	200.3	(3.5)
reg.F.Brab	279.8	(5.6)	262.9	(3.6)
reg.F.East	261.9	(5.5)	245.7	(3.6)
reg.F.Limb	277.7	(5.5)	263.2	(3.6)
reg.F.West	270.1	(5.5)	262.3	(3.6)
sex.Male	−13.5	(3.7)	−9.6	(2.1)
age[a]	0.75	(0.18)	0.27	(0.10)
length[b]			−88.6	(2.0)
position[c]			−0.004	(0.001)
lag.1			0.025	(0.002)
lag.2			0.008	(0.002)
lag.3			0.006	(0.002)
lag.4			0.006	(0.002)
lag.5			0.005	(0.002)
Random				
$\sigma^2_{u_{0j}}$ [d]	506.1	(388.1,636.9)	529.9	(421.8,668.3)
$\sigma^2_{u_{\mathrm{length}\ 0j}}$			595.1	(472.1,744.5)
$\sigma_{u_{0j}}\sigma_{u_{\mathrm{length}\ 0j}}$			−469.9	(−602.8,−364.1)
$\sigma^2_{e_{ij}}$ [e]	7564.6	(7473.1,7656.1)		
$\sigma^2_{e_{\mathrm{N}ij}}$			1396	(1365.0,1429.3)
$\sigma^2_{e_{\mathrm{F}ij}}$			3440	(3370.4,3511.7)
$\sigma^2_{e_{\mathrm{length}\ ij}}$			3424	(3312.3,3524.5)
$\sigma_{e_{\mathrm{N}ij}}\sigma_{e_{\mathrm{length}\ ij}}$			−1917	(1963.5,1870.5)
$\sigma_{e_{\mathrm{F}ij}}\sigma_{e_{\mathrm{length}\ ij}}$			−3044	(3113.0,2967.4)
Deviance	623 952		546 270	

[a]Speaker's age in years, centralized, in ms/year.
[b]Length of phrase in syllables, log-transformed and centralized, in ms/log(syllables).
[c]Sequential number of phrase within interview, centralized, in ms/number.
[d]$\sigma^2_{u_{0j}}$ denotes the variance between the j higher-level units (speakers).
[e]$\sigma^2_{e_{ij}}$ denotes the variance between the i lower-level units (phrases) within the j higher-level units (speakers).

about 1.3 times as long as those of the fastest speakers. For the present sample of 160 speakers, the comparable slowest 24 speakers yield a mean ASD of 186 ms, for the fastest 24 this is 297 ms. The present ratio of about 1.6 between slowest and fastest speakers' ASD is somewhat larger than the ratio 1.3 reported for vowel durations. (Comparing only speakers within each country yields ratios of 1.4 for both The Netherlands and Flanders). Finally, the preliminary model (1) confirms that within-speaker variance is indeed far larger than between-speaker variance in speech tempo (Miller *et al.*, 1984b), here by more than one order of magnitude. The subsequent model was developed to investigate this within-speaker variance further.

The preliminary model was extended by including several within-speaker predictors: *phrase length* (in syllables, converted to log units, and centralized to the mean log length), *sequential position* of each phrase within its interview (also centralized), and autocorrelation predictors with lags 1–5. (No multicollinearity was observed among the latter five predictors, with $r < 0.2$ for all pairs of predictors, and showing low eigenvalues for their principal components.)

Contrary to ANOVA models, multilevel models do not require the assumption that random variances (between speakers and within speakers) are homogeneous. Instead, such variances are modeled explicitly, which allows us to investigate the effects of the predictors on these variance components (Luke 2004; Snijders and Bosker, 1999; Pinheiro and Bates, 2000; Quené and Van den Bergh, 2004). The optimal model, specified in Eq. (2), contains effects of phrase length in its random parts (in parentheses) at both levels, and of country at the level of phrases within speakers. Speakers (higher-level units) vary in their average speech tempo. In addition, speakers may vary in their individual slopes of the phrase length effect. Between-speaker variance components are pooled over the speaker groups. Phrases within speakers (lower-level units) vary in their average speech tempo, and this variance within speakers may be different (nonhomogeneous) for speakers from The Netherlands and from Flanders. In addition, phrases within speakers vary in their

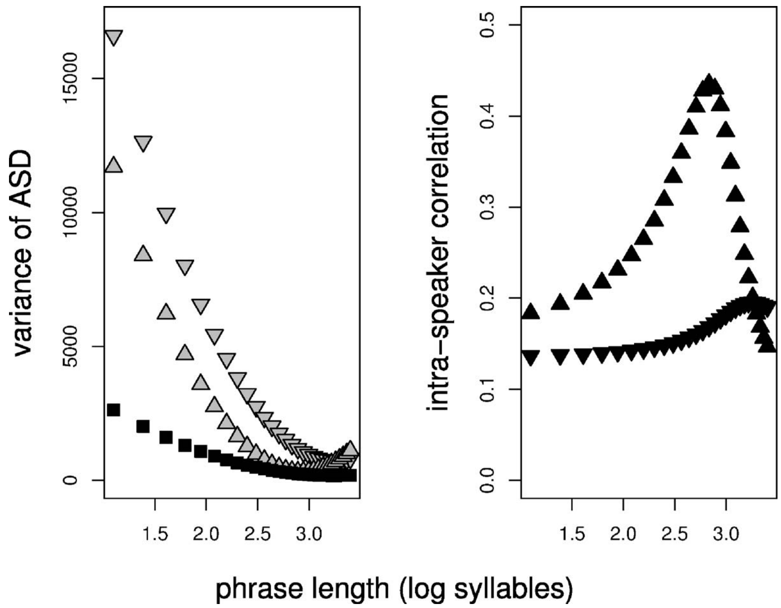

FIG. 1. Left panel: Variance estimates broken down by phrase length (log syllables), for between-speaker (dark squares) and within-speaker (light triangles) variances in ASD. Right panel: Intraspeaker correlations $[\sigma_u^2/(\sigma_u^2+\sigma_e^2)]$ broken down by phrase length (log syllables). Both panels show separate patterns for speakers from The Netherlands (triangles) and Flanders (inverted triangles).

slopes of the phrase length effect, and again this variance in slope may be different for the two speaker groups.

$$
\begin{aligned}
y_{ij} = {} & \text{reg.NR}[\gamma_{\text{NR }00}] + \text{reg.NM}[\gamma_{\text{NM }00}] \\
& + \text{reg.NN}[\gamma_{\text{NN }00}] + \text{reg.NS}[\gamma_{\text{NS }00}] \\
& + \text{reg.FB}[\gamma_{\text{FB }00}] + \text{reg.FE}[\gamma_{\text{FE }00}] \\
& + \text{reg.FL}[\gamma_{\text{FL }00}] + \text{reg.FW}[\gamma_{\text{FW }00}] \\
& + \text{sex.Male}[\gamma_{\text{sex }00}] + \text{age}[\gamma_{\text{age }00}] \\
& + \text{length}[\gamma_{\text{length }00} + (u_{\text{length }0j}) + (e_{\text{length }ij})] \\
& + \text{position}[\gamma_{\text{position }00}] + \text{lag.1}[\gamma_{\text{lag.1 }00}] \\
& + \text{lag.2}[\gamma_{\text{lag.2 }00}] + \text{lag.3}[\gamma_{\text{lag.3 }00}] \\
& + \text{lag.4}[\gamma_{\text{lag.4 }00}] + \text{lag.5}[\gamma_{\text{lag.5 }00}] \\
& + (u_{0j} + \text{N}[e_{\text{N }ij}] + \text{F}[e_{\text{F }ij}]). \qquad (2)
\end{aligned}
$$

Results for this model are listed in the right-hand part of Table I.

First, results for the *fixed* part confirm that phrase length indeed has a large and highly significant effect on speaking rate (Z=44.3, $p<0.001$), as known from previous research [Nooteboom (1972); Lindblom and Rapp (1973); De Rooij (1979); Nakatani *et al.* (1981)]. Speakers produce longer phrases with shorter average syllable duration, hence with faster speech tempo. In addition, sequential position has a small effect on speaking rate. Speakers tend to speed up by a very small amount (yielding somewhat shorter syllable durations) during the interview.

Second, one would expect that the effects of the between-speaker predictors (sex, age, country, and region) will be reduced in magnitude in model (2), where within-speaker variation is taken into account. The effect of sex (Z=−4.7, $p<0.001$) is indeed smaller than in the preliminary model, and it is below the 5% JND for speech tempo (Quené, 2007). The effect of age also decreases (Z=2.71, p=0.003). The reduction of the sex and age effects, and the significant phrase length effect, together suggest that the significant effects of sex and age observed in previous analyses may have been inflated by between-speaker effects on phrase length. This issue will be further discussed in the following. The tempo difference between speakers from The Netherlands (mean ASD 213 ms) and Flanders (mean ASD 263 ms) is again highly significant [$\chi^2(1)$=886.1, $p<0.001$]. Tempo was faster for speakers from the Randstad region (the linguistic center of The Netherlands) than for speakers from other regions of The Netherlands, although the difference is not significant [$\chi^2(3)$=6.6, p=0.083]. Regional differences within Flanders are significant in the current model (2) [$\chi^2(3)$=26.0, $p<0.001$], but speakers from the Brabant region (the linguistic center of Flanders) did not produce the fastest speaking rates.

Third, the positive autocorrelation coefficients suggest that speaking rate is weakly correlated among subsequent phrases. Phrases that are spoken faster (e.g., because of their pragmatic content) tend to be followed by somewhat faster phrases. These autocorrelation effects weaken even more as distance increases.

The *random* part of model (2) shows several interesting effects, regarding the effect of phrase length and the country effect (differences between The Netherlands and Flanders). These effects are illustrated in Fig. 1 for phrase lengths from 3 to 30 syllables, for which data are available from both countries (this includes 98% of phrases). First, within-speaker variances are smaller for speakers from The Netherlands (triangles pointing North) than for those from Flanders (triangles pointing South), as shown in the left-hand panel. As before, the within-speaker variance is considerably greater than the between-speaker variation in average or habitual speaking rate.

Second, phrase length affects the variances between speakers (dark symbols). As phrase length increases, individual speakers tend to converge to the same (fast) speaking rate, with decreasing variation between speakers. This decreasing variance presumably reflects universal phonetic constraints. In order to produce many syllables in one breath, speakers need to attain a fast tempo. For these long phrases, with associated fast tempo, speakers' average tempo is limited more by universal physiological and phonetic constraints (which reduce between-speaker variance) than by speakers' individual properties (which enlarge variance between speakers).

Even more striking than the decrement in between-speaker variance, however, is the similar decrement in within-speaker variance in ASD (Fig. 1, left-hand panel, triangles). As phrase length increases, phrases tend to converge to the speakers' average (fast) tempo for phrases of that particular length, with decreasing variation between phrases within speakers. The variance between phrases is presumably due to the semantic and pragmatic properties of each phrase (emphasis, attitude, etc.), as well as to its phonetic properties (phonetic complexity of the constituent syllables, phonological vowel length, etc.).

A large difference is observed between speakers from The Netherlands (triangles pointing North) and Flanders (triangles pointing South) in their within-speaker variances. Speakers from Flanders vary their tempo more (around the speaker's average) than speakers from The Netherlands do. This difference is probably not due to universal phonetic constraints, which presumably apply equally to speakers from The Netherlands and Flanders. Speakers apparently have some individual freedom in how strongly they vary tempo between phrases. The pattern in Fig. 1 suggests a cultural difference in within-speaker variations in tempo. The pattern also suggests that the (unknown) semantic-pragmatic effects are smaller for longer phrases.

In multilevel modeling, the intraspeaker correlation expresses the relative "uniqueness" of the higher-level units (speakers). This intraspeaker correlation is computed from the amounts of variances between and within speakers. The intraspeaker correlation is high if most variance is found between speakers (i.e., observations within an individual speaker are highly correlated) and only a small amount of variance is observed within speakers between phrases. The right-hand panel of Fig. 1 shows this intraspeaker correlation for the two speaker groups, as a function of phrase length.

For shorter phrases (up to about 20 syllables), the speakers from The Netherlands have a higher individual "uniqueness" than those from Flanders, mainly due to the lower within-speaker variances of the former, which makes their tempo more predictable. Speakers from The Netherlands adhere more to their own habitual average tempo than those from Flanders, yielding a higher intraspeaker correlation for the former. Stated otherwise, speakers from Flanders produce more expressive variation in tempo, which makes their tempo less predictable from the individual speaker's average, yielding a lower intraspeaker correlation.

For longer phrases, however, which are more challenging, the individual "uniqueness" of the speakers from The Netherlands decreases suddenly. All speakers from this group produce these phrases at the same tempo, with little variation between speakers, and with little (semantic-pragmatic) expressive variation within speakers. The remaining variance is presumably mainly due to universal phonetic constraints of the produced phrases. The pattern is different for speakers from Flanders, who still produce expressive variation even for these long and challenging phrases (as shown by their larger within-speaker variances). Speakers from Flanders also have to let go of their expressive variation, as phrase length increases, but they do so only for the longest phrases, of about 27 syllables or more. This ties in with their overall lower speech tempo. Because their average speech tempo is slower, speakers from Flanders still have some room for tempo variation even for very long phrases. Speakers from The Netherlands, by contrast, by speaking faster on average, are closer to their phonetic limitations, and have little room for variation in very long phrases.

In summary, these results suggest that speech tempo is slower, and has more expressive variation, for speakers from Flanders as compared to those from The Netherlands. Results also confirm that speech tempo increases with phrase length. Interestingly, average speech tempo is *not* affected by a speaker's age, if tempo measurements are corrected for phrase length, as in model (2). Without such corrections, however, age appeared to yield a significant main effect on speech tempo [Verhoeven *et al.* (2004) and preliminary model (1)]. The most likely explanation for this discrepancy between the models is that phrase length, as an intermediate variable controlling speech tempo, is itself affected by the speaker's age. For the sake of argument, let us assume that a speaker's mean phrase length is influenced by his or her age. In turn, phrase length affects speech tempo, so that speakers differing in (mean) phrase length would yield artifactual differences in speech tempo, if phrase length is not taken into account. If phrase length is included as a predictor for speech tempo, however, then the between-speaker differences in tempo would vanish, as was indeed observed in model (2).

This tentative explanation predicts that phrase length is indeed affected by the speaker's age (and perhaps also by other between-speaker predictors: sex, country, and region), in a way that is congruent with the observed differences in speech tempo. This was investigated by multilevel modeling of phrase length as the dependent variable.

B. Phrase length

Phrase length was also modeled by means of multilevel analysis, with speakers and phrases within speakers as two nested random factors. The optimal model, specified in Eq. (3), contains all remaining predictors in the fixed part, as well as age in the random part at the phrases-within-speakers level.

$$
\begin{aligned}
y_{ij} = {} & \text{reg.NR}[\gamma_{\text{NR }00}] + \text{reg.NM}[\gamma_{\text{NM }00}] \\
& + \text{reg.NN}[\gamma_{\text{NN }00}] + \text{reg.NS}[\gamma_{\text{NS }00}] \\
& + \text{reg.FB}[\gamma_{\text{FB }00}] + \text{reg.FE}[\gamma_{\text{FE }00}] \\
& + \text{reg.FL}[\gamma_{\text{FL }00}] + \text{reg.FW}[\gamma_{\text{FW }00}] \\
& + \text{sex.Male}[\gamma_{\text{sex }00}] + \text{age}[\gamma_{\text{age }00} + (u_{\text{age }0j}) \\
& + (e_{\text{age }ij})] + \text{position}[\gamma_{\text{position }00}] + \text{lag.1}[\gamma_{\text{lag.1 }00}] \\
& + \text{lag.2}[\gamma_{\text{lag.2 }00}] + \text{lag.3}[\gamma_{\text{lag.3 }00}] \\
& + \text{lag.4}[\gamma_{\text{lag.4 }00}] + \text{lag.5}[\gamma_{\text{lag.5 }00}] \\
& + (u_{0j} + \text{N}[e_{\text{N }ij}] + \text{F}[e_{\text{F }ij}]). \qquad (3)
\end{aligned}
$$

Resulting estimates for this model are listed in Table II.

Results for the *fixed* part of this model suggests that speakers from The Netherlands produce shorter phrases than those from Flanders, yielding a significant main effect of country on phrase length [$\chi^2(1)=56.0$, $p<0.001$]. Second, the phrase length is similar for speakers from different regions within the Netherlands [$\chi^2(3)=3.10$, n.s.] and within Flanders [$\chi^2(3)=5.54$, n.s.], yielding a nonsignificant main effect of region within country [$\chi^2(6)=8.64$, n.s.]. Third, phrases from female speakers are equally long as those from male speakers, as indicated by the insignificant coefficient for the sex factor. Fourth, the main effect of age is significant ($Z=2.08$, $p=0.019$), which indicates a tendency for older speakers to produce shorter phrases than younger speakers do. The log of phrase length decreases by 0.0027 for each 1 year increment of age.

TABLE II. Estimated parameters of multilevel modeling of phrase length (in syllables, log-transformed and centralized). Estimates of fixed parameters are given with standard error (in parentheses); estimates of random parameters are given with 95% confidence intervals obtained from Markov chain Monte Carlo sampling (in parentheses).

	Model (3)	
Fixed		
reg.N.Randstad	−0.012	(0.041)
reg.N.Mid	−0.049	(0.041)
reg.N.North	−0.0002	(0.041)
reg.N.South	−0.081	(0.041)
reg.F.Brab	0.195	(0.043)
reg.F.East	0.094	(0.042)
reg.F.Limb	0.153	(0.042)
reg.F.West	0.204	(0.043)
sex.Male	−0.017	(0.026)
age[a]	−0.0027	(0.0013)
position[b]	0.000	(0.000)
lag.1	0.000	(0.000)
lag.2	0.000	(0.000)
lag.3	0.000	(0.000)
lag.4	0.000	(0.000)
lag.5	0.000	(0.000)
Random		
$\sigma^2_{u_{0j}}$ [c]	0.027	(0.021,0.034)
$\sigma^2_{u_{\text{age }0j}}$	0.000	n/a
$\sigma_{u_{0j}}\sigma_{u_{\text{age }0j}}$	0.000	n/a
$\sigma^2_{e_{\text{N }ij}}$ [d]	0.209	(0.204,0.214)
$\sigma^2_{e_{\text{F }ij}}$	0.335	(0.328,0.343)
$\sigma^2_{e_{\text{age }ij}}$	0.00005	(0.00002,0.00008)
$\sigma_{e_{\text{N}ij}}\sigma_{e_{\text{age }ij}}$	−0.00008	(−0.00024, 0.00009)
$\sigma_{e_{\text{F}ij}}\sigma_{e_{\text{age }ij}}$	0.00059	(0.00031,0.00091)
Deviance	79 489.8	

[a]Speaker's age in years, centralized, in ms/year.
[b]Sequential number of phrase within interview, centralized, in ms/number.
[c]$\sigma^2_{u_{0j}}$ denotes the variance between the j higher-level units (speakers).
[d]$\sigma^2_{e_{ij}}$ denotes the variance between the i lower-level units (phrase) within the j higher-level units (speakers).

The *random* part of model (3) shows no effects of speaker's age on the between-speaker variance in phrase length. [If the random part is simplified to a single term σ^2_u at the higher level, ignoring age as a between-speaker predictor, then that simpler model fits the data equally well. Model (3) is preferred here, however, because its parametrization corresponds more closely to the above mentioned model (2) of the tempo data, allowing comparable analyses.]

The within-speaker effects of age and of country (differences between The Netherlands and Flanders) are illustrated in Fig. 2. First, within-speaker variances are again smaller for speakers from the Netherlands (triangles pointing North) than for those from Flanders (triangles pointing South), as shown in the left-hand panel. Not surprisingly, within-speaker variance in phrase length is considerably greater than the between-speaker variation (squares).

Second, the effect of age on within-speaker variance is small, but highly significant. [If the within-speaker effect of age is removed from model (3), then the log-likelihood increases to 79509.3, $\chi^2(3)=19.5$, $p<0.001$]. Figure 2 illustrates this age effect. The left-hand panel shows the amount of between-speaker variance (squares) and within-speaker variance (triangles). Within-speaker variance in phrase length decreases with age for young adult speakers, and then gradually increases with age for older adult speakers. The turning point seems to lie around age 35 for Flemish speakers, and around age 45 for speakers from The Netherlands. For the speakers over 45, this age effect on within-speaker variance is stronger for those from Flanders than from The Netherlands, as indicated by the steeper slope of the former. The right-hand panel of Fig. 2 shows the amount of intraspeaker correlation, i.e., speaker's "uniqueness" or individuality, across the age range. As before, the lower within-speaker variation for the speakers from The Netherlands corresponds with higher intraspeaker correlations for this group.

IV. GENERAL DISCUSSION AND CONCLUSION

The first aim of this study was to model between-speaker and within-speaker effects on speech tempo, and on tempo variations. The results for the full model (2) in Table I show two robust effects in its fixed part. First, tempo differs considerably between The Netherlands (mean 213 ms) and Flanders (mean 263 ms). Second, tempo is strongly affected by the length of the phrase, with anticipatory shortening as the connecting mechanism. There are also small effects of speaker's region, sex, and age. Tempo variation within

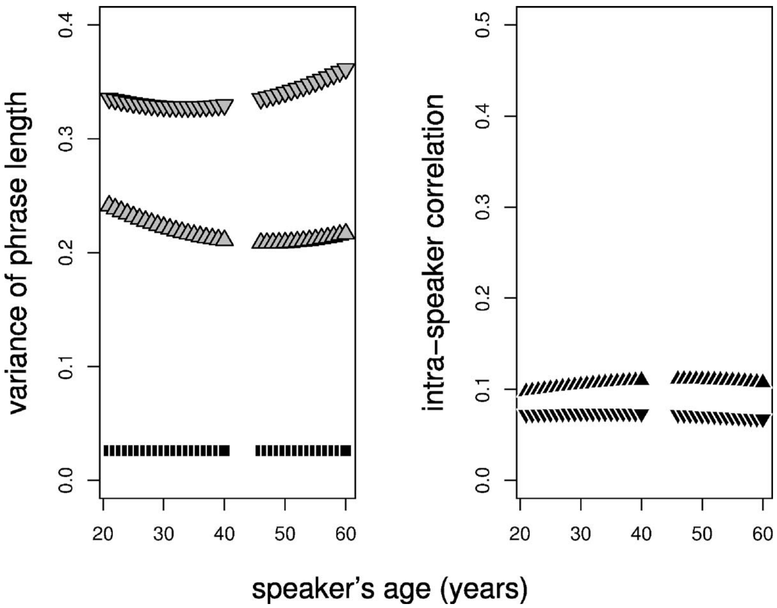

FIG. 2. Left panel: Variance estimates broken down by speaker's age (in years), for between-speaker (dark squares) and within-speaker (light triangles) variances in phrase length (which was log-transformed and centralized, see the text). Right panel: Intraspeaker correlations $[\sigma_u^2/(\sigma_u^2+\sigma_e^2)]$ broken down by speaker's age (in years). Both panels show separate patterns for speakers from The Netherlands (triangles) and Flanders (inverted triangles).

speakers is considerably larger than between speakers, but less so for speakers from The Netherlands than from Flanders. The faster tempo of speakers from The Netherlands leaves them less room for expressive variation, i.e., for within-speaker variations in tempo, presumably because these speakers are closer to their phonetic limitations. If they would speak slower, they might run out of breath before the end of the phrase; if they would speak faster, they might run into articulatory and perceptual difficulties.

Speech tempo in a phrase strongly depends of the length (in log syllables) of that phrase. This phenomenon is known as anticipatory shortening (Nooteboom, 1972; Lindblom and Rapp, 1973; De Rooij, 1979; Nakatani *et al.*, 1981): speakers shorten their syllables if they anticipate more syllables within a phrase. This finding also supports previous findings for American English read speech (Crystal and House, 1990). In the latter study, ASD was predicted in single-level fashion by seven phrase-internal predictors (e.g., proportion of stressed syllables, number of phones, etc.). This yielded an $R^2=0.579$. Since all six speakers read the same text, ASDs could be correlated between phrases spoken by pairs of speakers, yielding correlations between 0.66 and 0.88. ASD depends for a large part on the contents of a phrase [Crystal and House (1990), p.111], although it is also sensitive to other factors. In the present study, spontaneous speech instead of read speech was investigated. The spontaneous nature will probably lead to greater tempo variability within speakers, i.e., to a lower R^2 than observed for read speech (Crystal and House, 1990), because of the larger variation in pragmatic and semantic properties among phrases.

Previous analyses of speakers' aggregated tempi have suggested that dialect region, sex, and age are significant predictors of this between-speaker variation [Verhoeven *et al.* (2004) and preliminary model (1)]. The present study, however, nuances these findings. First, the differences among regions within each country may be significant, but these differences are not robust. The different rank orders of the regions' average tempi in models (1) and (2) and Verhoeven *et al.* (2004) constitute a warning against overinterpretation of these regional differences. Second, the sex difference may be significant, with males speaking somewhat faster than females, but this small difference is below the JND for speech tempo. Third, speech tempo is only weakly affected by speaker's age, if phrase length is also taken into account.

One possible explanation for these different findings in the present study is that the previously reported effects of age (and perhaps of sex, too) may have been indirect consequences of systematic variation in phrase length. If older speakers produce relatively shorter phrases than younger speakers, then this difference in phrase length would explain the observed age effect in speech tempo if phrase length is ignored, as well as the absence of an age effect in tempo if phrase length is not ignored. This predicted pattern was indeed observed for one between-speaker predictor, viz., speaker's age. In other words, between-speaker effects of age are mainly attributed to between-speaker differences in phrase length, with anticipatory shortening as the causal mechanism.

Phrase length turns out to vary with the speaker's age, in two ways. First, as predicted, older speakers produce shorter phrases than younger speakers do, as shown by the negative regression coefficient for the age predictor. Second, the amount of within-speaker *variation* in phrase length varies with the speaker's age (in years). Speakers over 45 tend to vary their phrase length more as the speaker's age increases. From age 45 to 60, within-speaker variance of (log-transformed) phrase length increases by about 4% (The Netherlands) to 8% (Flanders). Although modest in scale, the effect of age on within-speaker variance in phrase length is highly significant. The age grading effects may be explained by two mechanisms. Older speakers may successfully vary phrase length for communicative purposes, after decades of experience in expressing themselves as teachers. On the other hand, older teachers may increasingly suffer from cognitive constraints, both in retrieving words from their mental lexicon (e.g., Burke and Shafto, 2004) and in sentence construction (e.g., Kemper *et al.*, 2003). These cognitive constraints may have hampered the older speakers in this study more than the younger speakers, forcing the former to produce shorter phrases occasionally, and yielding more variation in phrase length. More research is needed to further investigate such possible explanations.

Phrase length was also predicted to vary with speaker's sex, but this was not observed in the present corpus study. Hence the somewhat faster tempo observed for males cannot be explained phonetically by their longer phrases (through anticipatory shortening), and cannot be explained by physical sex differences (e.g., in mass of articulators), which tend to be in the opposite direction. The small but significant differences in tempo are therefore most likely coupled to gender differences in the speakers' social dominance and status. This tentative explanation is supported by the tendency for male interviewees to produce more syllables during the whole interview (average 4210 syllables, $s=1093$) than female interviewees do [average 3926 syllables, $s=922$; $t(158)=1.777$, $p=0.077$]. Male interviewees are more talkative than female ones in the present corpus. Similar gender differences in talking behavior have been reported for a large corpus of telephone conversations, where male speakers produced more words than female speakers [mean 926 words versus 867 words, respectively, in mixed-sex conversations; Liberman (2006)], as well as for formal meetings, where male participants talk longer, and interrupt more often, than female participants (Holmes, 1995). This gender difference is also reported in a recent meta-analytical review (Leaper and Ayres 2007; cf. Mehl *et al.*, 2007). In addition to producing more words and syllables, and interrupting more often, male speakers may also express their social dominance by speaking somewhat faster than female speakers.

Phrase length also differs between speakers from The Netherlands and from Flanders, but not in the direction one would predict from the observed tempo difference between the countries. Speakers from The Netherlands produce significantly *shorter* phrases, spoken at a significantly faster tempo, than speakers from Flanders do. At this international level, then, the relation between phrase length and tempo is reversed. In addition, within-speaker variation in phrase length is again smaller in The Netherlands than in Flanders. These findings point toward a cultural difference in speaking styles between The Netherlands (allegro: shorter, faster, less varied) and Flanders (andante: longer, slower, more varied). Now that speech corpora are becoming available for other languages with multiple linguistic centers (French, Portuguese, English, etc.), it would be interesting to further investigate these intercultural differences in speaking styles, and their consequences for intercultural speech perception.

The present study also aimed at illustrating how multilevel modeling may help us understand variation at multiple hierarchical levels simultaneously (Quené and Van den Bergh, 2004). It allows us to model not only average effects, but also variances around these effects. This allowed us to observe, for example, that speakers from The Netherlands produce less within-speaker variation in their speech tempo than speakers from Flanders, yielding higher intraspeaker correlations for the former (Fig. 1). It also allowed us to observe that a speaker's age affects not only his or her *average* phrase length, but also the within-speaker *variation* in phrase length (Fig. 2). At present, multilevel modeling is the best statistical technique available for drawing such multiple inferences simultaneously. Moreover, it requires fewer assumptions, and allows for missing data. These properties make multilevel modeling perfectly suited for understanding data from semispontaneous speech corpora as used in the present study.

In the final multilevel model of speech tempo (2), most within-speaker variance may be attributed to phrase length and other predictors. But there still remains some unexplained within-speaker variance (of 1396 and 3440 variance units, for The Netherlands and Flanders, respectively). This suggests that other unknown factors, *not* related to phrase length or any other predictor, also control a speaker's expressive variations in speech tempo. In particular, emphasis and emotional involvement are known to affect speech tempo, as discussed earlier. These semantic, pragmatic, and affective factors cannot be investigated with the present corpus material. Most of the within-speaker variance in tempo, however, can be ascribed to the length of a phrase, due to anticipatory shortening. From the perspective of a listener, who has to use speech tempo as a scaling factor during speech perception, this is good news: most variation in tempo is predictable from an easily observed property of the speech stimulus.

In conclusion, the present study has yielded several robust estimates of variations in speech tempo, based on a large corpus of spoken Dutch. Between-speaker variations in habitual speech tempo are mainly due to the speaker's country background. Most variations in speaking rate occur within-speakers, however, and they are mainly related to the length of a phrase, due to anticipatory shortening. Speakers from The Netherlands produce faster speech, and with less expressive variation, than speakers from Flanders. Interestingly, phrase length itself decreases with speaker's age; this may have inflated previously reported effects of speaker's age on speech tempo. In addition, older speakers tend to produce more individual variation in phrase length; this may be due to a decrement in cognitive performance, or to an increment in their expressive verbal variation.

ACKNOWLEDGMENTS

Preliminary versions of this work were presented in a Festschrift *On Speech and Language: Studies for Sieb G. Nooteboom* (Utrecht: Netherlands Graduate School of Linguistics LOT, 2004), and at the Ninth European Conference on Speech Communication and Technology (InterSpeech), Lisbon, Portugal, 4–8 September 2005. My sincere thanks are due to Hans Van de Velde for his assistance in the corpus analysis, to Griet Depoorter (TST-Centrale) for help with speaker identification codes, to Huub van den Bergh for statistical guidance, and to Esther Janse, Sieb Nooteboom, Willemijn Heeren, Matthias Mehl, and two anonymous reviewers for helpful comments and suggestions.

Adank, P., Van Hout, R., and Van de Velde, H. (**2007**). “An acoustic description of the vowels of northern and southern Standard Dutch. II. Regional varieties,” J. Acoust. Soc. Am. **121**, 1130–1141.

Browne, W. (**2005**). *MCMC Estimation in MLwiN* (Centre of Multilevel Modeling, Bristol, UK.).

Burke, D. M., and Shafto, M. A. (**2004**). “Aging and language production,” Curr. Dir. Psychol. Sci. **13**, 21–24.

Cnaan, A., Laird, N. M., and Slasor, P. (**1997**). “Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data,” Stat. Med. **16**, 2349–2380.

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.

De Rooij, J. (**1979**). “Speech punctuation: An acoustic and perceptual study of some aspects of speech prosody in Dutch,” Ph.D. thesis, Rijksuniversiteit Utrecht, Utrecht, The Netherlands.

Goldman-Eisler, F. (**1968**). *Psycholinguistics: Experiments in Spontaneous Speech* (Academic, London).

Goldstein, H. (**1995**). *Multilevel Statistical Models*, 2nd ed. (Arnold, London).

Holmes, J. (**1995**). *Women, Men and Politeness* (Longman, London).

Hox, J. (**1995**). *Applied Multilevel Analysis* (TT Publicaties, Amsterdam).

Jakobson, R., and Waugh, L. R. (**1979**). *The Sound Shape of Language* (Harvester, Brighon).

Kemper, S., Herman, R., and Lian, C. (**2003**). “Age differences in sentence production,” J. Gerontol. B Psychol. Sci. Soc. Sci. **58B**, 260–268.

Koreman, J. (**2006**). “Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech,” J. Acoust. Soc. Am. **119**, 582–596.

Kreft, I. G., and De Leeuw, J. (**1998**). *Introducing Multilevel Modeling* (Sage, London).

Leaper, C., and Ayres, M. M. (**2007**). “A meta-analytic review of gender variations in adults’ language use: Talkativeness, affiliate speech, and assertive speech,” Personality and Social Psychology Review **11**, 328–363.

Liberman, M. (**2006**). “Gabby guys: The effect size,” Language Log Vol. **3607**, www.languagelog.org, last accessed 10 October 2006.

Lindblom, B., and Rapp, K. (**1973**). “Some temporal regularities of spoken Swedish,” Papers in Linguistics from the University of Stockholm, Vol. **21**, pp. 1–59.

Luke, D. A. (**2004**). *Multilevel Modeling* (Sage, Thousand Oaks, CA).

Mehl, M. R., Vazire, S., Ranirez-Esparza, N., Slatcher, R. B., and Pennebaker, J. W. (**2007**). “Are women really more talkative than men?” Science **317**, 82.

Miller, J., Aibel, I., and Green, K. (**1984a**). “On the nature of rate-dependent processing during phonetic perception,” Percept. Psychophys. **35**, 5–15.

Miller, J., and Grosjean, F. (**1981**). “How the components of speaking rate influence perception of phonetic segments,” J. Exp. Psychol. Hum. Percept. Perform. **7**, 208–215.

Miller, J., Grosjean, F., and Lomanto, C. (**1984b**). “Articulation rate and its variability in spontaneous speech: A reanalysis and some implications,” Phonetica **41**, 215–225.

Nakatani, L. H., O’Connor, K. D., and Aston, C. H. (**1981**). “Prosodic aspects of American English speech rhythm,” Phonetica **38**, 84–105.

Nooteboom, S. (**1972**). “Production and perception of vowel duration: A study of durational properties of vowels in Dutch,” Ph.D. thesis, Rijksuniversiteit Utrecht, Utrecht, The Netherlands.

Nooteboom, S., and Eefting, W. (**1994**). “Evidence for the adaptive nature of speech on the phrase level and below,” Phonetica **51**, 92–98.

Oostdijk, N. (**2000**). “Het corpus Gesproken Nederlands,” (The Spoken Dutch Corpus), Nederlandse Taalkunde **5**, 280–284.

Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed-Effects Models in S and S-Plus*, Statistics and Computing (Springer, New York).

Quené, H. (**2007**). “On the just noticeable difference for tempo in spontaneous speech,” J. Phonetics **35**, 353–362.

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.

Snijders, T., and Bosker, R. (**1999**). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage, London).

Stetson, R. (**1951**). *Motor Phonetics: A Study of Speech Movements in Action*, 2nd ed. (North Holland, Amsterdam).

Traunmüller, H., and Krull, D. (**2003**). “The effect of local speaking rate on the perception of quantity in Estonian,” Phonetica **60**, 187–207.

Tsao, Y.-C., Weismer, G., and Iqbal, K. (**2006**). “The effect of intertalker speech rate variation on acoustic vowel space,” J. Acoust. Soc. Am. **119**, 1074–1082.

Van Hout, R., De Schutter, G., De Crom, E., Huinck, W., Kloots, H., and Van de Velde, H. (**1999**). “De uitspraak van het Standaard-Nederlands: Variatie en varianten in Vlaanderen en Nederland” [The pronunciation of Standard Dutch: Variation and variants in Flanders and the Netherlands], in *Artikelen van de Derde Sociolinguïstische Conferentie*, edited by E. Huls and B. Weltens (Eburon, Delft, The Netherlands), pp. 183–196.

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.

Volaitis, L., and Miller, J. (**1992**). “Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories,” J. Acoust. Soc. Am. **92**, 723–735.

Winer, B. (**1971**). *Statistical Principles in Experimental Design*, 2nd ed. (McGraw-Hill, New York).

Zwaardemaker, H., and Eijkman, L. (**1928**). *Leerboek der Phonetiek: Inzonderheid met betrekking tot het Standaard-Nederlandsch* [Textbook of Phonetics: Particularly in relation to Standard Dutch] (Bohn, Haarlem, The Netherlands).