Speech rate effects on the processing of conversational speech across the adult life span[a)]

Xaver Koch[b)] and Esther Janse[c)]

Center for Language Studies, Radboud University, Erasmusplein 1, 6525 HT, Nijmegen, The Netherlands

This study investigates the effect of speech rate on spoken word recognition across the adult life span. Contrary to previous studies, conversational materials with a natural variation in speech rate were used rather than lab-recorded stimuli that are subsequently artificially time-compressed. It was investigated whether older adults' speech recognition is more adversely affected by increased speech rate compared to younger and middle-aged adults, and which individual listener characteristics (e.g., hearing, fluid cognitive processing ability) predict the size of the speech rate effect on recognition performance. In an eye-tracking experiment, participants indicated with a mouse-click which visually presented words they recognized in a conversational fragment. Click response times, gaze, and pupil size data were analyzed. As expected, click response times and gaze behavior were affected by speech rate, indicating that word recognition is more difficult if speech rate is faster. Contrary to earlier findings, increased speech rate affected the age groups to the same extent. Fluid cognitive processing ability predicted general recognition performance, but did not modulate the speech rate effect. These findings emphasize that earlier results of age by speech rate interactions mainly obtained with artificially speeded materials may not generalize to speech rate variation as encountered in conversational speech. © *2016 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4944032]

[VB]   Pages: 1618–1636

I. INTRODUCTION

Older adults, particularly those who are hearing impaired, report that they face challenges in speech comprehension in adverse listening conditions, such as when there is background noise or talkers have accents, mumble, speak softly or rapidly. The effect of increased speech rate on older adults' speech comprehension performance has often been operationalized by using artificial time compression, which may approximate some of the difficulties reported with fast speech (e.g., Wingfield, 1996; Vaughan *et al.*, 2006). Several studies have shown that artificially time-compressed speech makes comprehension and recall more difficult than normal-rate speech, and that this speech rate effect is larger for older, compared to younger adults (Wingfield, 1996; Gordon-Salant and Fitzgibbons, 1999; but cf. Schneider *et al.*, 2005 and Gordon *et al.*, 2009). Furthermore, speech rate effects seem to interact with the linguistic characteristics of the presented stimuli. Wingfield *et al.* (2003) have found that for older adults increased speech rate made listening particularly challenging if the presented sentences were also syntactically complex.

Before we provide a more detailed account of the literature on this finding that the effect of increased speech rate is larger for older than younger adults (henceforth, the *age × speech rate* interaction), we raise the point that results obtained with artificial time compression may either underestimate or overestimate the difficulty that listeners experience with naturally produced fast speech. Schmitt and Moore (1989) compared comprehension performance for time-compressed versus naturally produced faster speech rate in older adults. Their results showed generally better comprehension scores for naturally speeded up or slowed down materials than for unselectively compressed/expanded speech, suggesting that artificial time compression presents a more difficult listening condition than naturally increased speech rate. In contrast, a recent study (Gordon-Salant *et al.*, 2014) has shown that the recognition of artificially time-compressed read sentences seems to over-estimate the recognition of natural fast-rate speech (see also Janse, 2004). Gordon-Salant and colleagues found that both younger and older adults showed better sentence recognition performance for artificially speeded speech (originally read at a normal rate) than for natural fast-rate sentences read aloud by a talker at a very fast rate. However, what may be crucial is whether instructing talkers to read out sentences at their ceiling rate (as in Gordon-Salant *et al.*, 2014 and Janse, 2004) is representative of rate variation as observed in conversational speech in which speakers themselves habitually speak or choose to speak at a particular rate. Unlike artificially time-compressed speech, instructing talkers to speak as fast as they can generally involves less clear articulation because

most speakers are only able to speed up their speech rate through reduction of segments and syllables. The present study aims to investigate how naturally varying speech rate, as encountered in conversational speech materials spoken by different speakers, affects listening performance in younger, middle-aged and older adult listeners.

We now return to the accounts that have been provided for the *age × speech rate* interaction finding (as observed with artificially speeded speech) introduced above. Several studies have provided explanations for this differential rate effect on older adults' comprehension or recall performance interaction (e.g., Wingfield *et al.*, 1999; Schneider *et al.*, 2005). A first account for older adults' problems with speeded speech is the "generalized slowing hypothesis," which is based on cognitive aging research (e.g., Cerella, 1990). Salthouse (1985, 1996) proposed that a reduction in processing speed leads to impairments in cognitive functioning ("processing-speed theory" of cognitive aging). A general slowing of brain functions in aging and thus a reduced processing speed will lead to comprehension problems if more information units are transmitted per unit of time than the processor can handle (Wingfield, 1996). Importantly, an individual's processing speed predicted the effect of speech rate on older listeners' performance in a study by Janse (2009) using artificially speeded speech. If domain-general slowing should be held responsible for older adults' problems with fast speech rates, then increased rates of visual text presentation can be expected to also differentially affect older adults, compared to younger adults. However, this was not the case in a study by Humes *et al.* (2007). In their study, effects of increased rate of visual presentation were similar for younger and older adults.

Age-related changes in hearing have been put forward as another possible explanation for the increased problems older adults may have with fast speech. Epidemiological data suggest that around 40% to 50% of the population aged between 50 and 90 years are affected by hearing decline defined as pure-tone average thresholds (averaged over 0.5, 1.0, 2.0, and 4 kHz) above 25 dB hearing level (HL) (Cruickshanks *et al.*, 1998). Hearing impairment and age were found to independently contribute to deficits in recognizing temporally manipulated speech (Gordon-Salant and Fitzgibbons, 1993).

A third account for the *age × speech rate* interaction is that auditory processing ability may be impaired in older adults. Thus, apart from a gradual decline in absolute hearing sensitivity particularly for the higher frequencies, aging is accompanied by problems with central hearing, such as changes in temporal processing (Fitzgibbons and Gordon-Salant, 2010). Relatedly, older adults' problems with fast speech have been linked to longer neural adaptation periods in older listeners. Longer adaptation processes in older adults, as evidenced by, e.g., higher gap detection thresholds in older than in younger adults (Gordon-Salant *et al.*, 2006; Pichora-Fuller *et al.*, 2006; Haubert and Pichora-Fuller, 1999), may negatively influence the perception of stop consonants in fast speech. In line with this auditory processing account, Schneider *et al.* (2005) argued that older adults process artificially time-compressed speech differently from younger listeners. Schneider and colleagues base their "perceptual hypothesis" on the "notion that older adults find it more difficult to handle speed-induced acoustic distortions than do younger adults" (Schneider *et al.*, 2005, p. 268), thereby arguing for age-related differences in sensitivity to signal manipulations, such as artificial time compression. Schneider *et al.* (2005) compared the effects of a linear type of time compression (eliminating every third amplitude sample, the sampling method) and a selective time compression method that particularly compresses steady-state segments and leaves rapid transitions intact. Indeed, the results of Schneider *et al.* (2005) showed that younger and older adult groups were equally affected by increased speech rate when speech was speeded in a way that produced minimal acoustic degradation.

More evidence for acoustic degradation induced by artificial time compression algorithms comes from Kusomoto and Vaughan (2004), who compared acoustic features of artificially speeded-up (Synchronous-OverLap-Add technique) and natural speech. Their results suggest that for higher compression rates durational cues for plosive and fricative consonants may differ from natural speech. As durational cues are exploited in speech perception (e.g., Klatt, 1976; Raphael and Dorman, 1980), artificial speeding techniques may complicate speech processing, particularly at higher compression rates. Thus, artificial time compression changes perceptually relevant durational cues, which impairs speech comprehension, and this effect may be more pronounced for older than younger listeners (e.g., Goy *et al.*, 2013).

In sum, studies on age and individual differences in the effect of speech rate on speech perception so far have mainly focused on artificially time-compressed speech. Moreover, most studies have focused on sentences that were read aloud. Importantly, Wingfield *et al.* (1999) state that recall of auditorily presented speech passages drops significantly if the presented speech rates exceed "normal limits" (p. 385), particularly for older adults. Gordon *et al.* (2009) also state that *age × speech rate* interaction effects usually occur if materials are speeded to rates beyond those found in normal speech.

This raises the question as to which speech rates can be considered "normal" and what is a "normal" range? Speech rate is operationalized as the number of linguistic units (e.g., words, syllables, phones) per unit of time (e.g., minute, second). In contrast to "articulation rate," "speech rate" includes pauses. Krause and Braida (2004) state that clear speech involves speech rates of about 100 words per minute (wpm, i.e., 2.3 syll./s)[1] and that conversational speech would easily involve a doubling of that tempo (i.e., 4.6 syll./s). The study of Greenberg (1998) of a spontaneous English discourse corpus showed a mean syllable duration of around 200 ms, i.e., an articulation rate of 5 syllables per second. For Dutch, Quené (2008) found a mean articulation rate of about 4.2 syllables per second in the interview part of the Spoken Dutch Corpus (Oostdijk, 2000). The unit of measurement in Quené (2008) was interpause chunks. The fastest speaker in this sample had a mean articulation rate of about 5.6 syllables per second and the slowest speaker a rate of 3.0

J. Acoust. Soc. Am. **139** (4), April 2016

Xaver Koch and Esther Janse   1619

syllables per second. The highest articulation rate Quené (2008) found in an interpause chunk was 12.1 syllables per second (Quené, 2014). In sum, a speech rate of about 4 to 6 syllables per second can be assumed typical for conversational speech in West Germanic languages such as English or Dutch. Speech rates roughly range between around 2 and 12 syllables per second. The *age × speech rate* interaction effect found by Janse (2009), for example, is based on the comparison of a rate that is 1.5 times normal rate (i.e., given that the normal rate in that study was 5.7 syllables per second, 1.5×5.7 syll./s $= 8.6$ syll./s) and a rate that was twice the normal rate (i.e., 2.0×5.7 syll./s $= 11.4$ syll./s). Both time-compressed conditions therefore, represent higher-than-typical speech rates. Speech rate studies have worked with higher-than-typical rates, and artificially speeding speech changes perceptually relevant durational cues (cf. Kusomoto and Vaughan, 2004). This raises the question whether experimental results obtained with artificial time compression generalize to processing of natural speech heard in everyday conversations. The present study therefore investigated how natural speech rate variation as found within and between speakers in a conversational speech corpus affects listening performance in adults of varying age (cf. Gordon et al., 2009).

As hypothesized by the perceptual and generalized slowing accounts of the *age × speech rate* interaction, the effect of speech rate on speech comprehension may interact with the listener's auditory, linguistic and cognitive abilities. We therefore included these participant-related variables into our modeling of perceptual performance. We investigated speech processing by employing the visual-world paradigm. This technique provides information on the time course of the recognition of a word embedded in a running sentence and yields complementary behavioral (click response times) and psychophysiological data (gaze data, pupil size data). Eye-tracking allows us to observe speech processing in real time as there "is no appreciated lag between what is fixated and what is processed" (Just and Carpenter, 1980, p. 331). The task-evoked pupil response reflects the cognitive demands of processing a stimulus (Zekveld et al., 2013). Speech rate is expected to affect ease of processing, and hence understanding faster stimuli is cognitively demanding. Cognitive demand affects the pupil response (e.g., Zekveld et al., 2013). We therefore hypothesized that processing effects that are related to increased speech rate should be reflected in click response times, gaze data, and in the task-evoked pupil response.

We address the following three research questions:

(1) Can we replicate speech rate effects on word recognition performance using conversational materials with naturally varying speech rates?
(2) Do younger adults, middle-aged adults, and older adults differ in the effect of speech rate on their word recognition performance?
(3) Which individual measures predict general word recognition performance and the effect of increased speech rate on recognition performance over the adult life span?

II. METHOD

A. Participants

Three age groups were included: older adults (aged over sixty years), middle-aged adults (between 30 and 60 years), and younger adults (between 18 and 30 years). None of the participants reported hearing difficulties. From the initial sample of 112 adults, 12 participants were excluded from the analyses for the following reasons. The semi-automatized eye-tracking calibration procedure was not successful for two participants (one older and one younger adult). The test session of one middle-aged participant was interrupted by construction noise. Furthermore, eight participants were excluded (seven older adults and one middle-aged) because hearing loss in one or both ears exceeded the Dutch prescription criterion for hearing aids [pure-tone average over 1, 2, and 4 kHz (PTA^{high}) > 35 dB HL]. One additional older adult was excluded because of very low task accuracy (less than nine percent of all 60 trials correct) while accuracy for the remaining participants ranged between 77% and 100% correct ($M = 97.1\%$, $SD = 3.3$, see Sec. II E). The final sample consisted of 100 Dutch participants, 32 older adults ($M^{age} = 67$ years, $SD = 4.7$, 20 females), 33 middle-aged adults ($M^{age} = 50$ years, $SD = 7.5$, 21 females) and 35 younger adults ($M^{age} = 21$ years, $SD = 2.5$, 22 females).

B. Background measures

Participants' hearing was screened in both ears with air conduction pure-tone audiometry using the Hughson-Westlake procedure (Carhart and Jerger, 1959) for octave frequencies from 0.25 to 8 kHz, including two half-octave frequencies of 3 and 6 kHz, see Fig. 1.

Audiometric thresholds for the better ear were entered as a covariate in our statistical modeling of word recognition performance. This was done as auditory presentation in the word recognition experiment was binaural: we assumed that hearing sensitivity in the better ear would at least partly compensate for hearing loss in the worse ear, such that taking the better ear, rather than the poorer ear, presents a conservative estimation of the effect of hearing loss on performance (cf. Chen et al., 2015). Four participants (one younger and three older adults) showed asymmetric hearing loss, defined as an interaural difference of more than 10 dB, averaged over 0.5, 1, 2, and 4 kHz (following Noble and Gatehouse, 2004). Table I lists descriptive and test statistics regarding the hearing sensitivity measures for the three age groups. Three different pure-tone average (PTA) measures were analyzed. (a) PTA^{low}: mean over 0.5, 1, and 2 kHz; (b) PTA^{high}: mean over 1, 2, and 4 kHz; and (c) high-frequency PTA (PTA^{HF}): mean over 3, 4, 6, and 8 kHz. Age groups particularly differed in the higher frequencies (cf. Table I for significant age group differences in PTA measures).

In addition to the assessment of hearing thresholds, all participants completed the following five tests: (a) a visual acuity test, (b) the Digit Symbol Substitution Test, (c) the vocabulary subpart of the Groningen Intelligence Test, (d) a
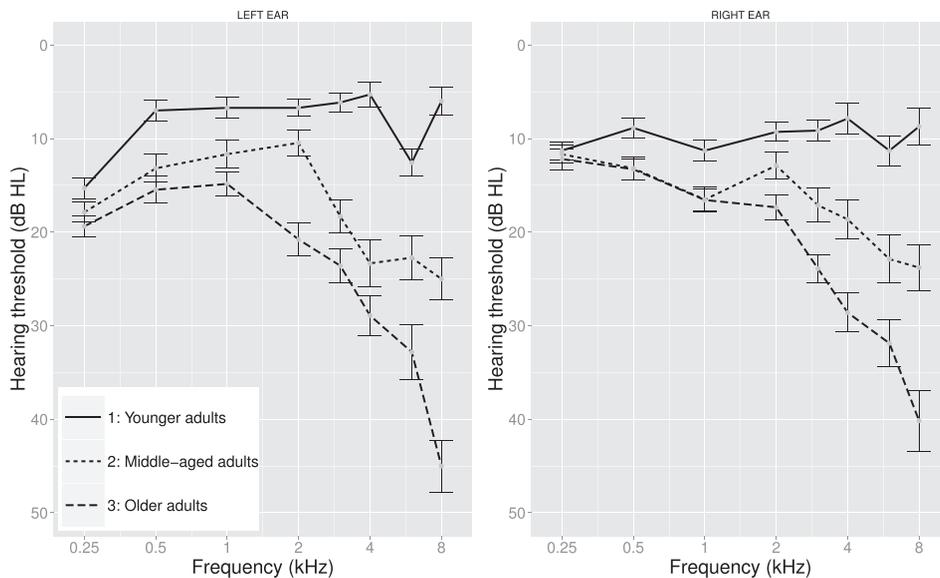
Xaver Koch and Esther Janse

FIG. 1. Mean audiometric pure-tone air conduction thresholds (for left and right ear) as a function of frequency for the younger, middle-aged and older adults. Error bars represent standard errors.

visual Digit Span Test with Backward recall, and (e) Raven's Standard Progressive Matrices Test. The five tests and the reasons for including them are described below.

1. Visual acuity test

Visual acuity was tested because all participants should be able to easily read the orthographic stimuli presented during the experiment (30 point Tahoma, i.e., approximately 0.8 cm height, see Sec. II D). Depending on whether participants wore their lenses or glasses during actual testing, their vision or corrected vision was tested to measure their (corrected) visual acuity. Acuity was assessed with the participant's head on a chinrest with constant 330 lux illumination. A standard Snellen visual acuity test chart was downscaled to be appropriate for the fixed test distance of 60 cm (being the fixed test distance during the eye-tracking experiment). Individual visual acuity was operationalized as the LogMAR equivalent (cf. Holladay, 1997) which is based on the logarithmic transformation of the Snellen fractions. Note that the LogMAR equivalent for normal vision is 0, with higher values representing poorer visual acuity. Mean visual acuity was 0.23 ($SD = 0.17$) and ranged between 0 and 0.57. Crucially, all participants were able to correctly read the row with the largest font on the test chart which was half as large as the orthographic stimuli presented during the experiment

(30 point Tahoma). As expected, visual acuity was poorer with higher age. All three age group comparisons showed significant age-related declines in visual acuity (cf. Table II).

2. Digit-Symbol-Substitution Test

Participants' individual processing speed was assessed with the Digit-Symbol-Substitution Test (DSST), which is a subpart to the Wechsler Adult Intelligence Test (Wechsler, 2004). Salthouse (2000) found that scores on the DSST relate to processing and perceptual speed. Importantly, DSST performance was included as it predicted how much the individual listener was impacted by increased speech rate (Janse, 2009). Test performance was operationalized as specified in the test manual (number of correctly re-coded items within two minutes). Processing speed generally declines with age (Salthouse, 2000), which is also evidenced in our data (cf. Table II).

3. Vocabulary test

The vocabulary subpart measure of the Groningen Intelligence Test (Luteijn and van der Ploeg, 1983) was included as an index of individual linguistic ability to investigate whether word recognition, and the effect of speech rate on word recognition, is associated with vocabulary size. During the computerized multiple-choice test participants

TABLE I. Means and standard deviations of pure tone average measures in the better ear for younger adults (YA), middle-aged adults (MA), and older adults (OA) and results of test statistics investigating age group differences in pure-tone average measures (Mann-Whitney-Wilcoxon, significance levels corrected for multiple testing).

Hearing variable	Age group			Comparisons		
	YA M(SD)	MA M(SD)	OA M(SD)	YA–MA p	YA–OA p	MA–OA p
PTA[low]	8.62 (4.55)	13.54 (6.51)	13.65 (4.05)	< 0.001	< 0.001	ns
PTA[high]	6.33 (4.53)	13.64 (6.71)	19.42 (5.48)	< 0.001	< 0.001	< 0.001
PTA[HF]	7.14 (5.29)	20.86 (11.40)	32.76 (12.61)	< 0.001	< 0.001	< 0.001

TABLE II. Means and standard deviations of non-auditory participant related variables for younger adults (YA), middle-aged adults (MA) and older adults (OA) and results of test statistics investigating age group differences (Mann-Whitney-Wilcoxon, significance levels corrected for multiple comparisons).

Background variable	Age group			Comparisons		
	YA $M(SD)$	MA $M(SD)$	OA $M(SD)$	YA–MA p	YA–OA p	MA–OA p
Visual acuity	0.10 (0.10)	0.23 (0.16)	0.37 (0.14)	0.002[a]	< 0.001	< 0.001
Processing speed	87.26 (13.46)	76.12(15.45)	64.56 (13.36)	0.018[b]	< 0.001	< 0.001
Vocabulary	13.83 (2.04)	15.79 (1.60)	16.63 (2.06)	< 0.001	< 0.001	ns
Working memory	55.95 (18.81)	63.64 (23.83)	48.96 (17.93)	ns	ns	0.012[b]
Reasoning	44.54 (5.60)	38.64 (5.99)	32.25 (8.20)	< 0.001	< 0.001	0.002[a]
Fluid cognitive processing ability	0.76 (0.68)	−0.01 (0.78)	−0.83 (0.83)	< 0.001	< 0.001	< 0.001

[a]Significance level notation: $p < 0.01$.
[b]$p < 0.05$.

had to select correct synonyms for 20 words (choice out of four options for each word). There was no time pressure to complete the test. Test performance was operationalized as the number of correct responses. Younger adults showed poorer vocabulary scores than middle-aged and older adults (cf. Table II).

4. Digit span test backwards

Many studies have shown that recognition of spoken sentences in noise is associated with individual working memory ability, verbal working memory in particular (e.g., Rönnberg et al., 2008, 2013). Furthermore, Small et al. (1997) demonstrated that individual working memory capacity modulates speech rate effects on speech comprehension. We selected a digit span test with backward recall to tap simultaneous storage and manipulation of verbal information. A computerized visual version of the Wechsler Adult Intelligence Test (Wechsler, 2004) digit-span test was administered. Participants had to recall 12 digit sequences after two practice trials. The digits in each sequence (two to seven items, increasing in length over trials) were presented one after another on a computer screen and participants were prompted to type in the digits in reverse order after presentation (digit-display time: 1000 ms, inter-stimulus interval: 200 ms). Individual performance was operationalized as the percentage of accurate trials. Middle-aged adults outperformed older adults in this task, but none of the other age group comparisons showed significant differences (cf. Table II).

5. Raven's standard progressive matrices test

A test of non-verbal reasoning was included to investigate whether non-verbal intelligence (as opposed to verbal abilities measured by digit span performance) relates to speech processing performance. A modified version of Raven's matrices test (Raven et al., 2003; henceforth, RAVEN) was administered in which a time limit was imposed to restrict the overall test session duration (cf. Wilhelm and Schulze, 2002). Participants were asked to complete as many items as possible within 10 min. Skipping items was prohibited. We modified the results

form and enlarged the font sizes to 14 point as the original version had a rather small font size (9 point). The RAVEN score reflects the sum of correct responses for all five matrices sets. The maximal score that could be obtained was 60 (5 sets × 12 items). The results in Table II show that reasoning abilities differ between the age groups with younger participants outperforming the middle-aged and older groups.

6. Correlations between background measures

We investigated possible intercorrelations between background measures and age using Spearman's rank-order correlation tests (cf. Table III). A moderate-to-strong correlation was observed between the nonverbal intelligence measure and processing speed (RAVEN and DSST, respectively, $r = 0.58$, $p < 0.001$) which may partly be due to a mental speed component in the speeded version of this reasoning task (cf. Wilhelm and Schulze, 2002). We ran a factor analysis using IBM SPSS Statistics 20 to derive a factor representing the common variance between the two cognitive measures. The use of this factor allowed us to avoid collinearity issues (redundancy) in our statistical modeling and enabled us to include a construct underlying both variables. The factor thus combines processing speed, which is linked to general (fluid) intelligence (e.g., Coyle et al., 2011), and reasoning abilities, which are thought to reflect general (non-verbal) intelligence. The analysis revealed an initial eigenvalue of the single factor explaining 79% of the variance with factor loadings of 0.89 both for processing speed and reasoning. Individual scores for each participant for the newly created composite variable "fluid cognitive processing ability" (cf. Park et al., 2010) were included in the statistical analyses.

C. Materials

1. Conversational stimuli

We specifically chose question-answer sequences (henceforth, QA sequences) for our test paradigm as these represent minimal conversational units which are "a reasonable proxy for turn-taking more generally" (Stivers et al., 2011, p. 10 588). Conversational fragments were selected

Xaver Koch and Esther Janse

TABLE III. Correlation matrix with correlation coefficients and significance levels for participant-related variables including age (Spearman's rank, significance levels corrected for multiple comparisons).

Background variable	Age	PTA^{low}	PTA^{high}	PTA^{HF}	Visual acuity	Processing speed	Vocabulary	Working memory	Reasoning
PTA^{low}	0.44[a]								
PTA^{high}	0.75[a]	0.59[a]							
PTA^{HF}	0.80[a]	0.42[a]	0.83[a]						
Visual acuity	0.64[a]	−0.38[b]	−0.48[a]	−0.59[a]					
Processing speed	−0.60[a]	−0.41[a]	−0.57[a]	−0.56[a]	−0.48[a]				
Vocabulary	0.50[a]	0.27	0.32[b]	0.34[b]	0.31[c]	−0.13			
Working memory	−0.13	0.04	−0.12	−0.24	−0.06	0.11	0.01		
Reasoning	−0.59[a]	−0.41[a]	−0.52[a]	−0.51[a]	−0.38[b]	0.58[a]	−0.19	−0.02	
Fluid cognitive processing ability	−0.67[a]	−0.46[a]	−0.61[a]	−0.60[a]	−0.48[a]	0.88[a]	−0.17	0.04	0.89[a]

[a]Significance level notation: $p < 0.001$.
[b]$p < 0.01$.
[c]$p < 0.05$.

from the spontaneous dialogue part (face-to-face component) of the Spoken Dutch Corpus (Oostdijk, 2000). The following three primary criteria were defined to extract stimuli from the corpus on the basis of the corpus' orthographic transcriptions and part-of-speech tagging: (a) the QA sequence had to consist of two speakers and one change of turns; (b) the minimal length of the question was two words (e.g., "Van wie?," "Of whom?"); and (c) the minimal length of the answer had to be five words (e.g., "Ik ga een zon maken," "I will make a sun"). The orthographic representations of the 1200 candidate QA sequences that met the criteria above were checked for coherence of question and answer by a native speaker of Dutch. Moreover, we selected QA sequences containing at least one Dutch mono- or disyllabic (trochaic) target noun in order to match syntactic category and length for the target words. To avoid prosodic boundary phenomena, we only chose question-answer sequences in which the target word was neither the first word nor the last word in the answer portion of the QA sequence. The resulting set of QA sequences was narrowed down further by excluding conversations with speaker overlaps and stimuli with loud background noise, as well as QA sequences containing pauses longer than 0.2 s in the answer part of the second speaker. Application of these criteria led to a set of about 90 short question-answer conversation fragments. Out of those 90 QA sequences, 60 instances were selected as target stimuli, plus a set of 15 filler QA sequences, showing the same conversational features as the targets. An example target QA sequence is given in the orthographic transcription below with the target word underlined in the Dutch transcription and English translation (all QA sequences are listed in the supplementary material[4]).

Example 1.
speaker 1: "*Waar was het nou toch?*"
　　　　　"Where was it again?"
speaker 2: "*Waar die ten hemel steeg*,"
　　　　　"Where he ascended to heaven."

Table IV shows the descriptive statistics of the variables related to the QA sequences used in our statistical analyses. Thirty-eight of the 60 target words were disyllabic, the remaining 22 target words were monosyllabic (monosyllabic structures varied in complexity from CVC to CCVCC;

disyllabic nouns varied in complexity from CV–CV to CCCVC–CVC, see the supplementary material[4]). We included the target words' number of syllables measure as a variable in our analyses as the uniqueness point for disyllabic target words may be earlier relative to word offset than for monosyllabic target words (see Sec. II E).

Target word duration ranged between 196 and 866 ms ($M = 372$ ms, $SD = 139$). Mean CELEX word frequency (Baayen *et al.*, 1993) for the 60 target words was 185 (occurrences per million tokens: English words having this frequency would be words such as *table, parents, evening, group*). Target word frequency values were logarithmically transformed to normalize their distribution ($M = 4.00$, $SD = 1.78$). Shapiro-Wilk normality tests showed that the log-transformed target word frequencies were normally distributed where the untransformed frequency values were not. Log-transformed target word frequency showed a statistically significant negative correlation with target word duration [Spearman's rank: $r(58) = -0.27$, $p < 0.05$]. This relation was expected on the basis of Zipf's law, which predicts more frequent words to be shorter (Zipf, 1949, 1965). Log-transformed target word frequency was

TABLE IV. Descriptives of the item-related variables used in statistical modeling.

Covariate	M	SD	Range
Speech rate (syllables/second)	5.91	1.80	2.93–11.22
Target word frequency (per 10^6 tokens, logtransformed)	4.00	1.78	0–7.22
Trial number (excl. 6 training trials)	45.13	21.40	7–81
Target word predictability	0.42	0.10	0.22–0.84
Target word position in the answer phrase	7.05	2.59	3–16
SNR (dB) for the answer phrase	23.79	5.34	12.43–37.42
Target word's number of syllables			1 | 2 $n = 22$: $n = 38$

included as a control variable in our analyses, as we expected more frequent target words to be easier to recognize (see Sec. II E).

The position of the target word in the answer phrase of the QA sequence ranged between the third and the sixteenth word ($M = 7.05$, $SD = 2.59$). We included target word position as an item-related (control) predictor in the analyses as it can be interpreted as a predictability measure. Our hypothesis was that having a later position in the sentence would facilitate target word recognition.

For each target stimulus, speech rate of the second speaker's answer fragment was calculated in syllables per second from answer onset until the end of the target word. We based this calculation on the canonical (dictionary-based) number of syllables for each word in the target passage, rather than on the number of realized syllables. Speech rates are normally distributed over the stimulus set, and ranged between 2.93 and 11.22 syllables per second ($M = 5.91$, $SD = 1.80$). Obviously, target word duration and speech rate of the test items were strongly negatively correlated [Spearman's rank: $r(58) = -0.57$, $p < 0.001$]. As an additional control covariate for our analyses, we also approximated the signal-to-noise ratio (SNR) for all items separately. For each item the background noise intensity level (noise floor for the channel of the target speaker recording) was subtracted from the mean intensity of the respective answer part of the target speaker ($M = 23.79$ dB, $SD = 5.34$). Speech rate and SNR were not correlated [Spearman's rank: $r(58) = 0.12$, $p > 0.1$].

To investigate how the spectral content of our speech materials compared to standardized materials, we compared the long-term average spectrum (LTAS) of our test stimuli to the LTAS of the International Speech Test Signal (Holube *et al.*, 2010) which, in turn, has been shown to be comparable to the international long-term average speech spectrum (Byrne *et al.*, 1994). This comparison did not show substantial differences between our 60 question-answer sequences and the ISTS material up to 4 kHz (mean difference 1 to 4 kHz over 100 Hz wide bins: 0.74 dB). We also checked whether the fragments with higher speech rate had a different spectral content from the lower speech rate fragments (by means of a mean split on speech rate). No systematic spectral differences were observed between the two sets of fragments.

The set of answer fragments in the 60 target stimuli involved 49 different speakers (age range: 19 to 76 years, $M^{age} = 37$ years, $SD = 18.7$). Eight target speakers were presented multiple times to the participants (maximally three times).

2. Orthographic stimuli

After the extraction of the target and filler QA sequences, orthographic stimuli were selected for the visual world paradigm employed in the eye-tracking experiment (McQueen and Viebahn, 2007). Three word categories were created for each target word: a semantic distractor, a phonetic distractor, and a phonetic distractor to the semantic distractor. The latter category was chosen to make the display symmetrical in that there were always two pairs of onset-overlapping words on the screen. The semantic distractors were derived from the same semantic field as the respective target words [e.g., for the target "hemel" ("sky") the semantic distractor "aarde" ("earth") was selected]. The phonetic distractors shared at least the initial phoneme with the respective target words but often also the following vowel [e.g., for the target "hemel" ("sky") the phonetic distractor "heling" ("handling stolen goods") was selected]. The phonetic distractor to the semantic distractor stimuli minimally shared the initial phoneme with the respective semantic distractors [e.g., for the semantic distractor "aarde" ("earth") the semantic-phonetic distractor "aanhef" ("salutation") was selected]. We verified that all orthographic distractors matched the morphosyntactic context of each individual sentence in terms of word class, number, and noun gender (as common gender nouns take a different definite article than neuter nouns in Dutch). For the 15 fillers the four orthographic stimulus categories were selected accordingly to ensure that participants could not tell upfront whether a stimulus was a target or a filler trial. The supplementary material[4] shows the set of distractors for the target words.

3. Assessing target predictability

All target QA sequences were tested for the predictability of the target word given the preceding conversational context. This was done in a separate test and allowed us to distinguish speech rate effects from effects of contextual target predictability (to be entered as control variable in our analyses). Note that this predictability measure differs from the position of the target word in the answer fragment introduced above: two target words that both occur as the fifth word in the phrase may still differ in how predictable they are given the prior words in the phrase (see correlations between measures below). Eighteen younger Dutch adults ($M^{age} = 19.6$ years, 14 female) participated in this predictability rating experiment. Participants were presented with orthographic representations of the 60 test QA sequences up until the target word and had to rate all four orthographic word stimuli (the target, the semantic distractor, the phonetic distractor, and the phonetic distractor to the semantic distractor) for their match with the given context on a scale from 0 (does not fit at all) to 100 (fits perfectly). Participants gave their ratings in a text processing program on a computer. Target word predictability scores were calculated for each QA sequence in two ways: as an absolute predictability score for the target word given the pre-context, and as a proportional value, being the target word's rating against the sum of ratings for all four orthographic representations (see the supplementary material[4]; descriptive statistics for the proportional values see Table IV). The proportional predictability score was calculated to take into account how probable the target word was relative to the predictability ratings of the three orthographic distractors (see Brouwer *et al.*, 2012). No significant correlations were found between target predictability and the following four item-related variables: speech rate in the answer fragment, target word frequency

(CELEX frequency), target word position in the answer phrase and SNR (none of the *r* values exceeded |0.20|, all Bonferroni-corrected *p* values exceeded 0.1).

D. Procedure

We set up a word recognition experiment, using the visual world paradigm (Allopenna *et al.*, 1998). On each experimental trial, participants had to click (with a computer mouse) the one out of four orthographically represented words they heard in a conversational speech fragment. We used a Desktop Mount Eyelink 1000 eye-tracker with a chinrest, a ViewSonic 22 in. screen monitor, circumaural Sennheiser HD 215 headphones plus a Hewlett Packard USB mouse. Testing took place in a sound-attenuating booth. Illumination was kept constant at 50 lux for all subjects during the whole test procedure to allow for pupillometry measurement (cf. Zekveld *et al.*, 2010). Before participating in the eye-tracking experiment, all participants underwent a visual acuity check in the test booth with a near vision Snellen chart (cf. Sec. II B 1). Participants first read the experiment instructions outside the test booth and could ask questions if anything was unclear. Instructions were repeated by displaying them on the computer screen after the eye-tracking calibration process just before the practice trials were presented. Each trial consisted of three phases: a talker familiarization phase, a preview phase and a response phase. During talker familiarization participants listened to short fragments of the two speakers they would hear in the upcoming QA sequence. These audio fragments consisted of about two seconds of speech (minimally a six syllable utterance), and were not related, content-wise, to the test QA sequence. Each of the familiarization fragments was preceded by an announcement spoken by a female speaker whom they would be hearing next ("speaker 1" or "speaker 2"). The order in which speakers were introduced matched the order of the speakers in the upcoming test stimulus. After this familiarization phase a fixation cross was presented for 300 ms centered on the screen. After talker familiarization and fixation cross, participants got a preview of the four (candidate) words on the screen for a period of three seconds. During this preview phase participants could read the four words silently to be prepared for the upcoming test conversation. These words (cf. Sec. II C 2 and the supplementary material[4] for more details) were presented in a black sans serif Tahoma font (30 point, bold letters) on a white background in four equal sized quadrants of a centralized section of the display. Apart from the four quadrant areas, an additional click region was present on the screen for both target and filler trials. This region (a centralized smaller grey colored circle labeled "none of these words") had to be clicked if none of the words on the screen had been perceived. After the three seconds preview period, listeners were presented with the question-answer sequence with the target word embedded in the second speaker's answer (for test trials) or without a target word (for filler trials). The familiarization and test stimuli were presented binaurally (same signal for both channels), at an intensity level of 70 dB sound pressure level using headphones.

The four words were displayed throughout the entire preview and response period until the participants clicked one of the words or the "none of these words" area on the screen. On each trial, the mouse cursor was reset to the screen center at the beginning of the preview phase. Each participant was presented with the same set of test and filler items. Participants were randomly assigned to one of four lists that comprised the 60 target stimuli plus 15 filler items in a different randomized order. Each of the word categories (target, phonetic distractor, semantic distractor, semantic-phonetic distractor) occurred equally often in each of the four quadrants on each of the randomization lists. Each participant got the same set of six practice trials before the experiment started (three target trials and three filler trials) to familiarize them with the task. Test duration for the eye-tracking procedure was approximately 22 min.

E. Analyses

Two types of models were set up: age group comparison models and individual-differences models (across age groups). Speech rate was the continuous within-subjects variable ($n = 60$) of interest in both types of models. Age group was entered as a between-subjects factor (younger adults, middle-aged adults, older adults) in the age group comparison models. For each of the investigated dependent variables (click response times, target word gaze probability, pupil dilation latency, pupil dilation amplitude) separate statistical regression models were run using linear mixed-effect models in the program R with the lme4 package (Bates *et al.*, 2013). As additional control variables (within-subjects predictors), we included the frequency of the target word (continuous), trial number (continuous), target word predictability (continuous), the target word's position in the answer phrase (continuous), SNR (continuous), and the target word's number of syllables (two levels: monosyllabic, disyllabic). Target word frequency (log-transformed) was included as an (item-related) control variable in our analyses because we hypothesized that more frequent targets would be easier to recognize. Trial number was included as control variable in the click response time (henceforth, click RT) and the pupil data analysis. Our hypothesis was that fatigue or practice effects would be covered by including trial number. To control for context effects, we included the target word's predictability in our modeling. We hypothesized that items with a higher predictability would be easier to recognize. As noted above, the target word's position in the answer phrase can also be interpreted as a predictability measure. It is however a more local context measure than the predictability measure above as it only covers the number of words prior to the target in the answer sequence of the second speaker. Our hypothesis was that more prior words would facilitate word recognition. We also hypothesized the SNR of the items to have an effect on performance. We expected word recognition to be easier at higher SNRs. We included the target word's number of syllables as a variable in the model as the uniqueness point (i.e., that point where the transcription of the word makes it unique relative to all

other words) of disyllabic target words may be earlier in the word than for monosyllabic words. Thus, if we measure from word offset, disyllabic words may show shorter click response times, faster looks to the target and faster pupil dilation responses than monosyllabic words.

Word recognition accuracy (as evidenced by clicking on the correct target) was investigated with generalized linear mixed-effect modelling (fixed effect: age group, random effects: participant, test item). Accuracy of the click responses was 98.1% for the younger adults, 98.3% for middle-aged adults, and 94.8% for the older adults. The analysis showed a significant age effect with older participants performing slightly worse than middle-aged $(B = -1.39, |z| = 4.76, p < 0.001)$[2] and than younger adults $(B = -1.24, |z| = 4.45, p < 0.001)$.

Only correct responses for the target trials were included in the subsequent data analyses (97.1% of all 6000 target trials). The models contained all item-related control variables as fixed effects (i.e., target word frequency, trial number, target word predictability, position of the target word in the answer phrase, SNR, number of syllables of the target word, see Table IV for descriptives). Trial number was included in the click response and the pupil data analysis. Trial number could not be included as a covariate into the gaze analysis as aggregated gaze data (over participants) did not contain information about the trial order anymore.

The age group comparison models also contained the critical interaction between the variables speech rate and age group. Individual-differences models also included participant-related variables (such as hearing, age in years, and scores on the five cognitive/linguistic tests) as fixed effects, as well as the critical interaction between speech rate and these participant-related variables.

The target word gaze probability data were analyzed with growth curve analysis (GCA) (Mirman, 2014) to capture the time course of participants' gaze behavior. GCA is a type of multilevel regression with which variation in curve shapes over time can be modeled. Thus, GCA can model linear and non-linear behavior of a dependent variable. In addition, a main advantage of GCA is that it does not involve alpha inflation due to repeated comparisons for multiple analysis windows. Given our GCA approach for the eye gaze data, the first, second and third order orthogonal time polynomials (linear, quadratic and cubic time component) were included in our gaze data analysis as fixed effects. The first orthogonal polynomial (Time[1]) describes linear change in the dependent variable over time, which is comparable to the slope in a linear regression model. The second time polynomial (Time[2]) captures the change of a dependent variable over time that follows a quadratic function (flat vs more bent curve). The third time polynomial (Time[3]) encompasses the curve shape as a product of a cubic time fitting function. A cubic function involves an additional twist in the curve compared to the quadratic function. In the age group comparison model, interactions between each of these three time variables and age group were included, as well as between each of the time variables and speech rate. Second, three-way interactions between age, speech rate and each of these three time variables were included.

Performance of the younger adults (i.e., the group mapped on the intercept) served as baseline for the age group comparison analyses. The random-effect structure of the models included random intercepts for participants and items where possible (i.e., for the click response time analysis and for the pupillometry data).

Due to the sparseness of the gaze data for GCA, we had to aggregate our data. For the age group comparison, we aggregated over participants within each age group. Consequently, the random structure of the gaze data age group comparison analysis had only item (and not participant) as a random intercept. No individual-differences model is reported here, as setting up such a model would have entailed aggregating over items (and hence over speech rate, being our variable of interest).

We allowed for the possibility that the effect of speech rate randomly differed across participants. We therefore added random slopes for speech rate to the random structure of our click response time and pupillometry data models. Additionally, for the gaze probability model, the orthogonal polynomial terms (time components) were added on the respective random intercept (test item). All continuous variables were z-transformed. As the linear mixed-effect models do not output significance levels, we obtained these test statistics by using the analysis of variance function of the R car package which calculates type II Wald χ^2 values. For models including age group (which is an ordinal categorical variable with three levels: younger, middle-aged, older), p values were obtained using the model's t values. The number of degrees of freedom was estimated via the formula $df = J - m - 1$ (Hox, 2010), with J being the most conservative number of second-level units (number of items in our study, $n = 60$) and m being the number of included predictors.

Below we will describe the dependent variables (click response time, target gaze probability, pupil dilation latency and pupil dilation amplitude) separately to elaborate on the necessary data transformation steps.

1. Click response time

A priori we expected increased speech rate to make word recognition more difficult such that click response times would be slower (cf. Janse, 2009). We measured click response times from target word offset such that we did not have to take word duration into account, which was correlated with speech rate. Outliers were removed per age group for all accurate trials: responses slower than 2.5 *SD*s above each respective age group's mean click response time (in ms) were excluded (148 trials). The exclusion of inaccurate and outlier responses resulted in 95% of the recorded click response times being fed into the analyses (95% of the data points for the younger, 96% for middle-aged adults and 93% for the older adults). Mean click response time after outlier removal was 1030 ms after word offset $(SD = 507)$. As expected, middle-aged and older adults generally gave slower responses than younger participants (younger adults: $M = 877$ ms; $SD = 444$; middle-aged adults: $M = 1026$ ms; $SD = 508$; older adults: $M = 1205$ ms; $SD = 515$).
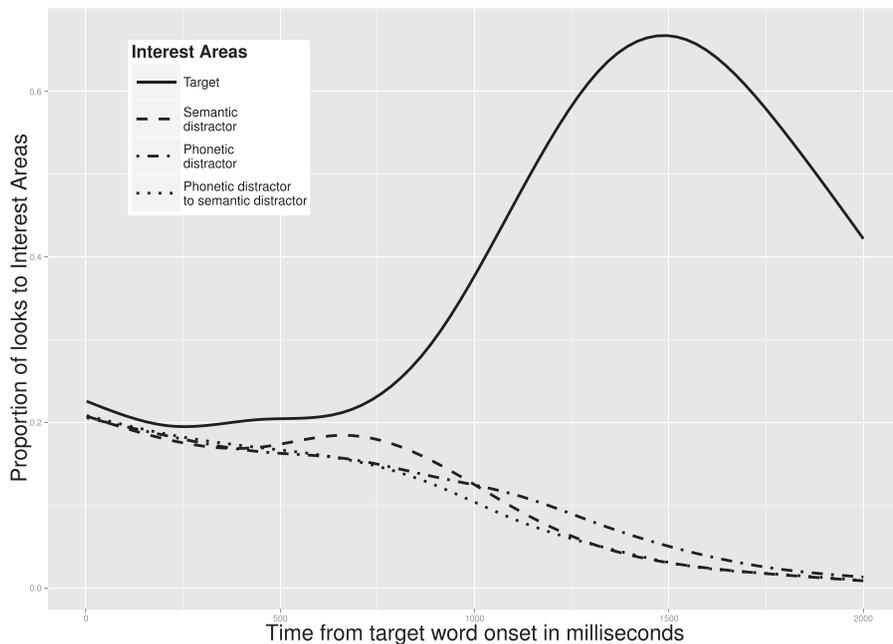
Xaver Koch and Esther Janse

FIG. 2. Grand mean of the fixation proportions to target and distractor words over time (measured from word onset).

2. Target gaze probability

We expected that increased speech rate would make word recognition more difficult, and that the increased word recognition difficulty would result in a lower probability of correct target gazes (cf. Ben-David *et al.*, 2011) and in a lower slope of the rise in target gaze probability in the analysis window.

Gaze fixations to the five interest areas (i.e., to the four word quadrants and to the "none of these words" area) were investigated in the time window between 200 and 1400 ms after target word onset. The onset of the analysis window was chosen because programming a saccade takes approximately 200 ms (cf. Barr, 2008). The window's upper limit was set to 1400 ms (given the mean click RT as measured from word onset).

Binomial gaze data for the interest areas were transformed to gaze probability on a log-odds scale (empirical logit, see Barr, 2008) over 24 consecutive time bins of 50 ms between 200 and 1400 ms after target word onset. As noted above, for the analysis of the *age × speech rate* interaction gaze data were aggregated over participants within each age group for each item (the items varying continuously in speech rate). Consequently, 1440 data points (24 time bins × 60 items/speech rates) were available per age group for GCA.

Mean probability (log odds ratio) of looking at the target over the 24 time windows for the three age groups was -0.85 (unit empirical logit, $SD = 1.10$, range: -7.47–3.07) which corresponds to a mean probability of 29.9% (range: 0.06%–95.6%). This probability was very similar across age groups (younger adults: $M = -0.88$, $SD = 1.06$, range: -7.47–1.67; middle-aged adults: $M = -0.79$, $SD = 1.10$, range: -7.38–2.47; older adults: $M = -0.87$, $SD = 1.12$, range: -7.38–3.07). Figure 2 shows the gaze probability curves for target and distractors from target word onset

for the analysis interval (200–1400 ms after target word onset).

3. Pupillometric data

Several pupil measures have been reported to reflect cognitive effort in language processing, such as mean pupil size and pupil peak latency (Andreassi, 2000; Zekveld *et al.*, 2013; Kuchinski *et al.*, 2013; Schmidtke, 2014). We investigated the task-evoked pupil response (pupil peak latency, pupil peak amplitude) for the word recognition task starting from target word onset. Our hypothesis was that faster speech rates would result in higher processing demands, yielding a delayed dilation response with a higher peak amplitude (cf. Beatty, 1982; Zekveld *et al.*, 2010). Pupil size data were recorded with a sampling rate of 500 Hz. For each trial the last 500 ms of the (silent) preview phase served as a baseline for the size of the pupil. This time interval was chosen for an item-individual baseline correction as visual input during baseline was the same as during listening in the test phase (having the 4 candidate words for this trial on the screen). Consequently, mean baseline pupil size could be subtracted from pupil size data points for the analysis interval. Trials with a high rate of missing values (i.e., more than 3.0 SDs above the mean) for the baseline interval were excluded from further processing (resulting in exclusion of 2.0% of all accurate trials, 117 trials). Missing values in the remaining baseline data were imputed by linear approximation (na.approx function, R package zoo). We then applied a locally weighted polynomial regression fitting algorithm in R (loess function, package stats, settings: span = 0.1, degree = 1). For each trial, we calculated a unique baseline pupil size value averaged over the fitted data in the baseline time window. For pupil size during the test window, the pupil size data in the time window between 500 ms before target word onset until the participant's click response was processed. Trials with a high incidence of

TABLE V. Click response time data (in ms): Model testing for the *age (group) × speech rate* interaction.

Fixed effects	Estimate	SE	|t|	p<
Intercept: Younger adults	934.97	49.60	18.85	
Middle-aged vs Younger adults	151.43	57.65	2.63	0.012[a]
Older adults vs Younger adults	328.30	57.97	5.66	0.001[b]
Speech rate: Younger adults	84.25	20.34	4.14	0.001[b]
Speech rate: Middle-aged vs Younger adults	−4.00	16.39	0.24	0.809
Speech rate: Older adults vs Younger adults	−1.08	16.04	0.07	0.947
Target word frequency	18.86	18.13	1.04	0.304
Trial number	−69.06	6.04	11.45	0.001[b]
Target word predictability	−4.82	19.48	0.25	0.806
Target word position	−16.53	18.39	0.90	0.373
SNR	15.64	17.67	0.89	0.381
Target word's number of syllables	−70.96	39.62	1.79	0.080[c]
Speech rate × trial number	23.91	6.14	3.89	0.001[b]

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.001$.
[c]$p < 0.1$.

missing values (i.e., more than 3.0 *SD*s above the mean) were excluded (resulting in exclusion of 1.1%, 65 trials). Missing values were imputed during the data smoothing and fitting procedure using a polynomial regression algorithm, which assigns less weight to outliers [smooth.m (Jiang, 2010) with method rloess, span = 0.2, MATLAB]. Baseline correction per trial was accomplished by subtracting the baseline mean pupil size value from each of the samples of the smoothed and fitted test data. The resulting baseline-corrected data showed a mean peak dilation maximum at around 1000 ms after target word onset which is in line with the timing of the canonical pupil response for processing a stimulus (cf. Zekveld *et al.*, 2010; Kuchinsky *et al.*, 2013). As a reminder, for the gaze data the analysis window was set between 200 and 1400 ms after target word onset (1400 ms being the mean click RT measured from word onset). However, based on the literature (Privitera *et al.*, 2008) and on visual inspection of the pupil response grand mean over all trials, we chose a different analysis window for the pupil dilation peak data. Privitera *et al.* (2008) report latencies of around 300 to 700 ms for the onset of the dilation phase. We therefore set our peak detection window between 500 and 1800 ms after target word onset (see Fig. 5 for the pupil dilation curves per age group). For each trial peak latency and peak amplitude were automatically extracted [peakdet.m (Lambrev, 2015), delta = 0.9, MATLAB]. Whenever there were multiple peaks in the detection window, the highest-amplitude peak was chosen. Automatic pupil peak detection was successful in 69% of the trials (with similar percentages of included trials for the different age groups). Mean pupil peak latency for the three age groups was 796 ms (*SD* = 266, range: 224–1380; younger adults: *M* = 756 ms, *SD* = 263; middle-aged adults: *M* = 792 ms, *SD* = 262; older adults: *M* = 842 ms, *SD* = 267). Mean pupil peak amplitude for the three age groups was 300.82 (arbitrary unit, *SD* = 200.80; younger adults: *M* = 351.08, *SD* = 232.10; middle-aged adults: *M* = 280.47, *SD* = 159.12; older adults: *M* = 270.26, *SD* = 194.11).
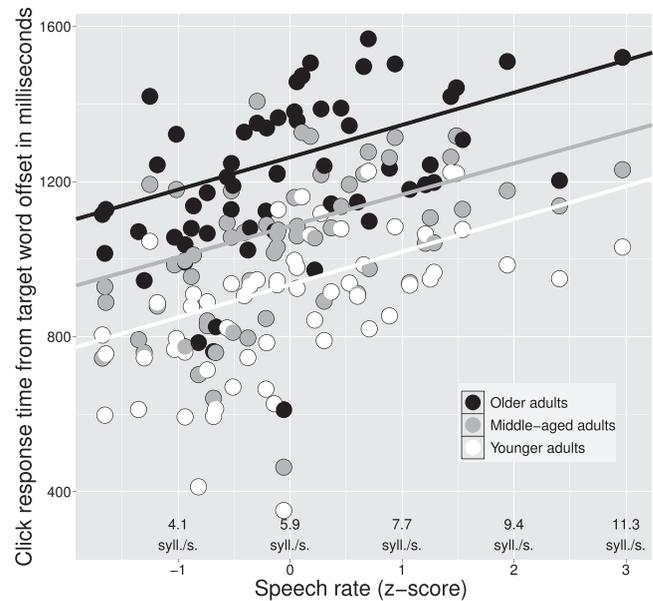

FIG. 3. Model predictions for click response times in ms per age group as a function of speech rate (speech rate-age group model). Points represent mean observed click response times per age group across speech rates (speech rate is given both as *z*-scores on the *x* axis, and for illustration purposes also as actual syllables per second). The model predictions (fit lines) take the contribution of all control variables into account.

F. Results

1. Click response time

The response time analysis was conducted to answer research questions 1 (Can we replicate speech rate effect using conversational speech materials?) and 2 (Do the three age groups differ in the effect of speech rate on their word identification performance?).

The result of the statistical model testing for the critical interaction between the predictors speech rate and age group (plus the control variables discussed in Sec. II E) is shown in Table V and is illustrated in Fig. 3.

Age groups differed significantly in their click response times ($t > 2.62$, $p < 0.02$, for both comparisons). As can be seen in the model in Table V and in the model plot in Fig. 3 younger adults showed the fastest click responses (approximately 930 ms at the mean speech rate of 5.9 syll./s) followed by middle-aged adults (approximately 150 ms slower than younger adults) with older adults having the slowest click RTs (approximately 330 ms slower than younger adults).

For younger adults (mapped on the intercept), the model predicted an increase in click response times of about 84 ms for an increase of one *z*-score in speech rate (cf. estimate for speech rate variable, $|t| = 4.14$, $p < 0.001$). This corresponds to an increase of 47 ms in click response time for an increase of one syllable per second and sums up to an effect size of around 390 ms for the tested range in speech rate (2.93–11.22 syll./s). Running the same model with the middle-aged group (rather than the younger group) mapped on the intercept, we observed a significant speech rate effect of 80 ms for an increase of one *z*-score in speech rate ($|t| = 3.59$, $p < 0.001$). Mapping older adults on the intercept resulted in a significant speech rate effect of 83 ms for an

Xaver Koch and Esther Janse

TABLE VI. Click response time data (in ms): Model testing for interactions between speech rate and participant-related variables.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept	1148.98	71.09	16.16	
Age	92.55	45.43	2.04	0.034[a]
Speech rate	83.31	20.01	4.16	0.001[b]
Target word frequency	19.83	18.33	1.08	0.280
Trial number	−69.40	6.04	11.50	0.001[b]
Target word predictability	−3.48	19.70	0.17	0.860
Target word position	−17.76	18.59	0.96	0.340
SNR	16.59	17.86	0.93	0.353
Target word's number of syllables	−64.95	40.07	1.62	0.106
Speech rate × trial number	23.87	6.14	3.89	0.001[b]
PTA[HF]	20.05	33.89	0.59	0.981
Visual acuity	16.73	29.14	0.57	0.473
Fluid cognitive processing ability	−74.53	30.39	2.45	0.003[c]
Vocabulary	−57.99	26.69	2.17	0.048[a]
Working memory	17.43	22.47	0.78	0.344
Speech rate × age	−1.26	12.86	0.09	0.923
Speech rate × PTA[HF]	−15.57	9.36	1.66	0.097[d]
Speech rate × visual acuity	2.56	8.03	0.32	0.750
Speech rate × fluid cognitive processing ability	−8.90	8.38	1.06	0.289
Speech rate × vocabulary	6.30	7.37	0.86	0.393
Speech rate × working memory	2.67	6.16	0.43	0.665
Trial number × age	8.08	11.04	0.73	0.465
Trial number × PTA[HF]	−13.91	8.25	1.69	0.092[d]
Trial number × visual acuity	5.72	7.10	0.81	0.421
Trial number × fluid cognitive processing ability	4.85	7.41	0.66	0.513
Trial number × vocabulary	−3.42	6.52	0.53	0.600
Trial number × working memory	10.55	5.44	1.94	0.053[d]

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.001$.
[c]$p < 0.01$.
[d]$p < 0.1$.

increase of one z-score in speech rate ($|t| = 3.53$, $p < 0.001$). As stated above, compared to younger adults, click response times of middle-aged and older adults were slower. However, the three age groups showed similar speech rate slopes, as evidenced by the insignificant age group by speech rate interaction estimates ($|t| < 1$, $p > 0.1$, for both comparisons; see Fig. 3 and Table V).

None of the (item-related) control variables, except trial number, had a significant effect on click response time (trial number effect: $|t| = 11.45$, $p < 0.001$). Click RTs decreased with increasing trial number (− 69 ms for each z score unit increase, i.e., 248 ms over the experiment), suggesting task familiarization. The model also showed a marginally significant effect of the target word being monosyllabic or disyllabic: as expected, disyllabic words tended to be recognized earlier relative to word offset than monosyllabic words (effect size: approximately 70 ms).

Our third question was which individual abilities would modulate the effect of speech rate on word recognition performance. This was investigated in a model testing for interactions between speech rate and all participant-related

predictors including chronological age (plus all control predictors related to item characteristics). We also tested for possible interactions between trial number and the participant-related variables to check for background variables that modulated the individual task familiarization effect of the participants. Table VI displays the resulting model.

As before, statistically significant effects of speech rate, age and trial number were observed. Slower click RTs were observed for items with higher speech rates ($|t| = 4.16$, $p < 0.001$) and for older compared to younger participants ($|t| = 2.04$, $p < 0.05$). Click RTs decreased over trials ($|t| = 11.50$, $p < 0.001$). None of the other control predictors affected click response time. Participants with better fluid cognitive processing ability and better vocabulary knowledge showed generally faster click RTs ($|t| > 2.17$, $p < 0.05$). Importantly, however, none of the participant-related variables showed significant interactions with speech rate. The variable trial number showed an interaction with speech rate such that speech rate effects became larger for later trials ($|t| = 3.89$, $p < 0.001$). This may relate to the general trial effect that participants speeded up their click responses over the experiment due to task familiarization. Possibly, stimulus-related effects, like speech-rate variation, become more apparent once response times are more closely time-locked to ongoing speech processing.

In sum, our click response time results confirmed that speech rate effects on word recognition performance can be found using conversational stimuli (research question 1). Second, the click response time data showed that the three age groups were equally affected by increased speech rate (research question 2). Concerning our third research question on individual differences in the effect of speech rate on word identification, none of the included cognitive, hearing-related or linguistic abilities was found to be associated with the size of the speech rate effect on click response times.

2. Target word gaze probability

As indicated in Sec. II E, GCA was used to analyze the time course of the target gaze data from 200 to 1400 ms after target word onset. As data aggregation was necessary for GCA, and the continuous variable speech rate was our variable of interest, we only carried out the age group comparison analysis. This analysis tested for the critical *age (group) × speech rate* interaction on the time course of the target word fixations (probability of looking at the target word). We hypothesized that higher speech rates would result in overall less fixations on the target word and in a shallower slope of the target fixation probability. Table VII provides the full resulting model.

The model showed a statistically significant effect of speech rate on the probability of looking at the target word ($\beta = -0.13$, $|t| = 3.42$, $p < 0.01$).[3] This means that speech rate affected the probability of fixating the target word, with higher speech rates leading to decreased target gaze probability.

With the middle-aged group on the intercept, the model outputs a speech-rate estimate of $\beta = -0.10$ ($|t| = 2.76$, $p < 0.01$) and with older adults as reference group the speech-rate β is -0.14 ($|t| = 3.68$, $p < 0.001$).

J. Acoust. Soc. Am. **139** (4), April 2016

Xaver Koch and Esther Janse   1629

TABLE VII. Target gaze probability data (empirical logit scale): GCA model testing for the *age (group) × speech rate* interaction over time.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept: Younger adults	−0.774	0.084	9.10	
Middle-aged vs younger adults	0.088	0.091	0.96	0.342
Older adults vs younger adults	0.007	0.091	0.07	0.941
Speech rate: Younger adults	−0.128	0.038	3.42	0.002[a]
Speech rate: Middle-aged vs younger adults	0.025	0.051	0.49	0.630
Speech rate: Older adults vs younger adults	−0.010	0.051	0.19	0.852
Target word frequency	0.028	0.040	0.70	0.490
Target word predictability	0.022	0.042	0.52	0.604
Target word position	0.086	0.040	2.18	0.035[b]
SNR	0.023	0.038	0.60	0.554
Target word's number of syllables	−0.165	0.086	1.93	0.060[c]
Time[1]	3.20	0.256	12.52	0.001[a]
Time[2]	1.30	0.183	7.11	0.001[a]
Time[3]	−0.178	0.146	1.21	0.232
Time[1] × Speech rate (younger adults)	−0.014	0.114	0.12	0.907
Time[1] × Middle-aged vs younger adults	0.336	0.277	1.21	0.231
Time[1] × Older adults vs younger adults	0.613	0.277	2.21	0.032[b]
Time[1] × Target word frequency	0.023	0.120	1.89	0.066[c]
Time[1] × Target word predictability	0.181	0.128	1.41	0.165
Time[1] × Target word position	0.117	0.120	0.98	0.334
Time[1] × SNR	0.079	0.117	0.68	0.500
Time[1] × Target word's number of syllables	0.018	0.259	0.68	0.500
Time[2] × Speech rate (younger adults)	0.167	0.082	2.05	0.047[b]
Time[2] × Middle-aged vs younger adults	0.295	0.198	1.49	0.144
Time[2] × Older adults vs younger adults	−0.148	0.198	0.75	0.460
Time[2] × Target word frequency	−0.124	0.086	1.44	0.157
Time[2] × Target word predictability	0.111	0.092	1.21	0.232
Time[2] × Target word position	−0.071	0.086	0.83	0.412
Time[2] × SNR	0.129	0.084	1.55	0.128
Time[2] × Target word's number of syllables	−0.240	0.186	1.29	0.203
Time[3] × Speech rate (younger adults)	−0.043	0.065	0.66	0.515
Time[3] × Middle-aged vs younger adults	0.237	0.159	1.49	0.143
Time[3] × Older adults vs younger adults	0.261	0.159	1.64	0.108
Time[3] × Target word frequency	−0.101	0.065	1.48	0.146
Time[3] × Target word predictability	−0.017	0.074	0.23	0.823
Time[3] × Target word position	−0.017	0.069	0.24	0.809
Time[3] × SNR	−0.095	0.067	1.42	0.163
Time[3] × Target word's number of syllables	−0.269	0.149	1.81	0.077[c]
Time[1] × Speech rate × middle-aged vs younger adults	0.088	0.155	0.57	0.573
Time[1] × Speech rate × older adults vs younger adults	0.228	0.155	1.47	0.149
Time[2] × Speech rate × middle-aged vs younger adults	−0.025	0.111	0.23	0.823
Time[2] × Speech rate × older adults vs younger adults	−0.156	0.111	1.40	0.168
Time[3] × Speech rate × middle-aged vs younger adults	−0.017	0.089	0.19	0.852
Time[3] × Speech rate × older adults vs younger adults	−0.006	0.089	0.07	0.947

[a]Significance level notation: $p < 0.001$.
[b]$p < 0.05$.
[c]$p < 0.1$.

We did not find age effects on target gaze probability ($|t| < 1$, $p > 0.1$, for both comparisons). However, the model shows that older adults have a higher linear increase (Time[1] component) of their target gaze probability over the analysis window ($\beta = 0.61$, $|t| = 2.21$, $p < 0.05$). As can be seen in Fig. 4, older adults differ in their target gaze behavior from the other two age groups mainly in the very first two to three time bins (i.e., 200–350 ms after target word onset). This steeper linear increase in gaze

probability may mainly be due to older adults' early gaze behavior (i.e., at the start of the analysis window). Note that target gaze probabilities were only around 10% for the older adults in the first two to three time windows (i.e., below the chance level of 20%).

While we did not find an effect of speech rate in interaction with the linear time term ($|t| < 1$, $p > 0.1$), the quadratic time term (Time[2] curvature) changed with increasing speech rate ($\beta = 0.17$, $|t| = 2.05$, $p < 0.05$). Thus, the higher the
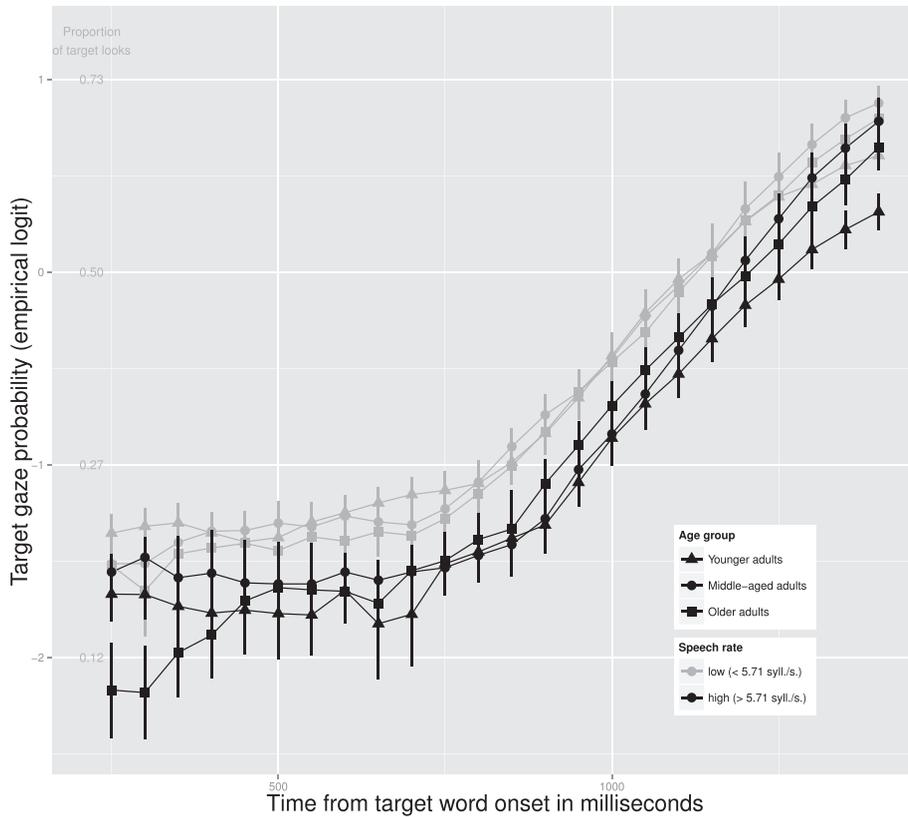
Xaver Koch and Esther Janse

FIG. 4. Target fixation probability over the analysis interval (200–1400 ms) for the three age groups for low and high speech rate items (median split on speech rate). Error bars represent standard errors.

speech rate, the more bent the gaze probability curve was, indicating a delayed target fixation pattern.

Whereas we observed both a speech rate effect (research question 1) as well as a general age (group) effect on target gaze probability (generally steeper linear increase for older adults), the model did not provide evidence for an *age × speech rate* interaction (research question 2). Figure 4 shows the gaze curves of the three age groups broken down

by speech rate (dichotomized for illustration purposes). Additionally, the model showed an effect of the control predictor target word position in the answer phrase: Items for which the target word was later in the answer phrase showed a higher probability of looks to the target ($\beta = 0.09$, $|t| = 2.18$, $p < 0.05$).

To conclude, our analysis of the time course of looking at the target word confirmed that speech rate effects on
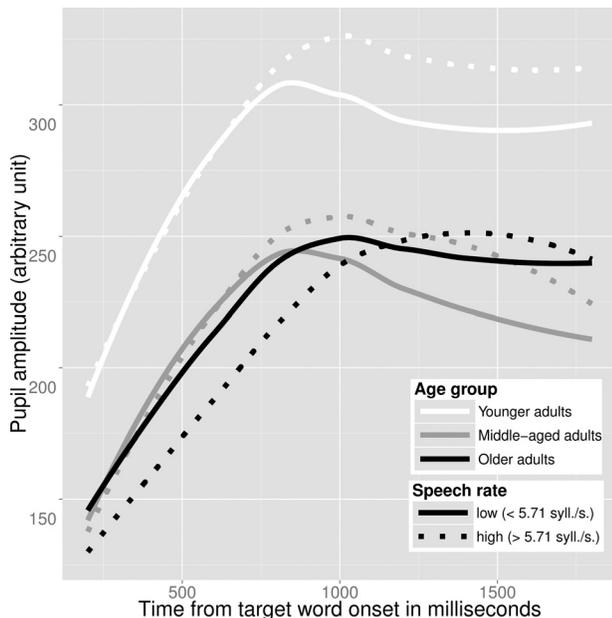

FIG. 5. Task evoked pupillometry response per age group for low and high speech rate items (median split on speech rate). The window chosen for peak detection was from 500 to 1800 ms after word onset.

TABLE VIII. Pupil peak latency data (in ms): Model testing for the *age (group) × speech rate* interaction.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept: younger adults	900.20	18.61	48.38	
Middle-aged vs younger adults	38.77	18.55	2.09	0.042[a]
Older adults vs younger adults	106.41	18.63	5.71	0.001[b]
Speech rate: Younger adults	5.04	10.66	0.47	0.639
Speech rate: Middle-aged vs younger adults	9.90	10.63	0.93	0.356
Speech rate: Older adults vs younger adults	3.01	10.57	0.29	0.777
Target word frequency	−11.84	8.43	1.41	0.167
Trial number	−17.90	4.86	3.68	0.001[b]
Target word predictability	−19.67	8.81	2.23	0.030[a]
Target word position	−17.78	8.39	2.12	0.040[a]
SNR	−14.64	8.09	1.81	0.077[c]
Target word's number of syllables	1.34	18.13	0.07	0.942
Speech rate × trial number	2.53	4.99	0.51	0.615

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.001$.
[c]$p < 0.1$.

TABLE IX. Pupil peak latency data (in ms): Model testing for interactions between speech rate and participant-related variables.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept	946.56	15.17	62.40	
Age	33.05	15.12	2.19	0.025[a]
Speech rate	9.17	8.73	1.05	0.272
Target word frequency	−12.09	8.43	1.43	0.152
Trial number	−17.97	4.85	3.70	0.001[b]
Target word predictability	−19.78	8.82	2.24	0.025[a]
Target word position	−17.55	8.40	2.09	0.037[a]
SNR	−14.73	8.10	1.82	0.069[c]
Target word's number of syllables	0.80	18.14	0.04	0.965
Speech rate × trial number	2.04	4.98	0.41	0.683
PTA[HF]	−2.80	11.16	0.25	0.818
Visual acuity	3.42	9.72	0.35	0.732
Fluid cognitive processing ability	−24.93	10.04	2.48	0.014[a]
Vocabulary	−8.78	8.88	0.99	0.319
Working memory	1.96	7.41	0.26	0.775
Speech rate × age	17.29	9.15	1.89	0.059[c]
Speech rate × PTA[HF]	−12.82	6.75	1.90	0.058[c]
Speech rate × visual acuity	−2.18	5.82	0.37	0.709
Speech rate × fluid cognitive processing ability	4.87	6.03	0.81	0.419
Speech rate × vocabulary	−1.36	5.29	0.26	0.797
Speech rate × working memory	6.97	4.40	1.58	0.114

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.001$.
[c]$p < 0.1$.

TABLE X. Pupil peak amplitude data (arbitrary unit): Model testing for the *age (group) × speech rate* interaction.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept: Younger adults	361.12	22.53	16.03	
Middle-aged vs younger adults	−79.46	31.02	2.56	0.014[a]
Older adults vs younger adults	−82.45	31.25	2.64	0.012[a]
Speech rate: Younger adults	0.34	5.40	0.06	0.951
Speech rate: Middle-aged vs younger adults	3.95	5.86	0.67	0.504
Speech rate: Older adults vs younger adults	0.66	5.82	0.11	0.911
Target word frequency	−10.46	4.06	2.58	0.013[a]
Trial number	−8.65	2.64	3.28	0.002[b]
Target word predictability	9.87	4.24	2.33	0.024[a]
Target word position	12.86	4.05	3.17	0.003[b]
SNR	−0.84	3.91	0.22	0.830
Target word's number of syllables	12.83	8.77	1.46	0.150
Speech rate × trial number	−2.12	2.72	0.78	0.440

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.01$.

word recognition performance can be found using conversational stimuli (research question 1). The gaze data also showed that the gaze behavior pattern of the three age groups was equally affected by increased speech rate (research question 2).

3. Pupillometric data

Two different analyses were conducted on the two pupillometry variables (pupil peak latency, pupil peak amplitude): one to address the age group comparison and the other to investigate individual differences. Figure 5 shows the time course of the pupil response per age group for low and high speech rates.

a. Pupil peak latency. The result of the statistical model testing for the critical interaction between the predictors speech rate and age group (plus the control variables discussed in Sec. II E) is shown in Table VIII.

Age groups differed significantly in their pupil peak latency ($|t| > 2.8$, $p < 0.05$, for both age group comparisons). Younger adults showed the fastest pupil peak dilation latency (approximately 900 ms at the mean speech rate of 5.9 syll./s) followed by middle-aged adults (approximately 39 ms slower than younger adults) with older adults having the slowest pupil peak dilation response (approximately 106 ms slower than younger adults). However, even though Fig. 5 suggests that pupil peak latency is affected by speech rate, the age group comparison model showed no significant speech rate effect on pupil peak latency ($|t| < 1$, $p > 0.1$). Furthermore, there was no evidence for a significant age group by speech rate interaction ($|t| < 1$, $p > 0.1$, for both comparisons; see Table VIII). Additionally, the model shows significant effects of target word predictability, target word position in the answer phrase and trial number. Pupil peak latency was smaller for more probable items and if the target word came later in the answer phrase ($|t| > 2.11$, $p < 0.05$, for both effects). Moreover, pupil dilation latency decreased over trials ($|t| = 3.68$, $p < 0.001$). Thus, all three described control variables facilitated word recognition.

A second model (see Table IX) was set up to investigate which individual abilities might modulate the effect of speech rate on pupil peak latency (note though that the pupil peak latency model above showed no speech rate effect). We tested for interactions between speech rate and all participant-related predictors including chronological age (and included all control predictors related to item characteristics).

Again, age showed a significant effect on the timing of the event-related pupil peak ($|t| = 2.19$, $p < 0.05$) with a slower pupil dilation response for older participants. Speech rate did not significantly affect pupil dilation latency ($|t| = 1.05$, $p > 0.1$). The only participant-related measure that significantly affected pupil peak latency (apart from age) was the composite factor fluid cognitive processing ability ($|t| = 2.48$, $p < 0.05$). Importantly, however, none of the participant-related variables showed significant interactions with speech rate. As in the age group comparison model above, the individual-differences model showed significant effects of target word predictability, target word position in the answer phrase and trial number. Pupil peak latency was shorter for more probable items ($|t| = 2.24$, $p < 0.05$) and the more words of the target speaker were available prior to the target word ($|t| = 2.09$, $p < 0.05$). Pupil latency also decreased over trials ($|t| = 3.70$,

Xaver Koch and Esther Janse

TABLE XI. Pupil peak amplitude data (arbitrary unit): Model testing for interactions between speech rate and participant-related variables.

Fixed effects	Estimate	SE	$|t|$	$p<$
Intercept	308.91	14.28	21.63	
Age	−24.37	26.66	0.91	0.364
Speech rate	1.72	4.22	0.41	0.684
Target word frequency	−10.47	4.07	2.57	0.011[a]
Trial number	−8.65	2.64	3.28	0.001[b]
Target word predictability	9.84	4.26	2.31	0.021[a]
Target word position	12.81	4.06	3.15	0.002[c]
SNR	−0.78	3.92	0.20	0.842
Target word's number of syllables	12.81	8.79	1.46	0.146
Speech rate × trial number	−2.03	2.73	0.75	0.456
PTA[HF]	−6.11	19.90	0.31	0.760
Visual acuity	2.51	17.18	0.15	0.883
Fluid cognitive processing ability	11.17	17.84	0.63	0.532
Vocabulary	−6.28	15.68	0.40	0.683
Working memory	−2.02	13.20	0.15	0.873
Speech rate × age	2.66	5.04	0.53	0.598
Speech rate × PTA[HF]	−1.33	3.72	0.36	0.721
Speech rate × visual acuity	0.46	3.21	0.14	0.887
Speech rate × fluid cognitive processing ability	−0.44	3.33	0.13	0.895
Speech rate × vocabulary	−2.35	2.92	0.80	0.422
Speech rate × working memory	−3.32	2.43	1.37	0.172

[a]Significance level notation: $p < 0.05$.
[b]$p < 0.001$.
[c]$p < 0.01$.

$p < 0.001$). In sum, contrary to the other dependent variables, our pupil peak dilation data did not show evidence that increased speech rate made spoken word recognition more difficult. Furthermore, none of the included participant-related variables was significantly associated with the size of the speech rate effect for the pupil peak latency.

b. Pupil peak amplitude. The result of the statistical model testing for the critical interaction between the predictors speech rate and age group (including the control variables discussed in Sec. II E) for pupil peak amplitude is shown in Table X.

Pupil peak amplitude differed considerably between the age groups ($|t| > 2.55$, $p < 0.05$, for both comparisons with younger adults mapped on the intercept). This is in line with earlier reports of reduced pupil size and less task-evoked pupil dilation for older participants (van Gerven *et al.*, 2004; Birren *et al.*, 1950). Yet, older adults did not differ significantly from middle-aged adults (as shown in a similar model with the middle-aged group on the intercept). The pupil peak amplitude model showed no simple speech rate effect ($|t| < 1$, $p > 0.1$), nor an *age group × speech rate* interaction ($|t| < 1$, $p > 0.1$ for both age group comparisons). As found for pupil peak latency, trial number affected pupil peak amplitude ($|t| = 3.28$, $p < 0.01$), suggesting task familiarization over the experimental trials. Additionally, the model showed significant effects of target word frequency, target word predictability

and target word position in the answer phrase. Unexpectedly, pupil peak amplitude was higher for more probable items ($|t| = 2.33$, $p < 0.05$) and for items that came later in the phrase ($|t| = 3.17$, $p < 0.01$). As expected, we observed a smaller pupil peak amplitude for words with a higher word frequency ($|t| = 2.58$, $p < 0.05$). A second pupil peak amplitude model was set up to test for interactions between speech rate and all participant-related predictors including chronological age (including all control predictors related to item characteristics; see Table XI).

In line with the age group analysis above, no speech rate effect was observed nor any significant interactions between speech rate and any of the participant-related variables. Similarly, consistent with the previous *age (group) × speech rate* model for the peak amplitude data, effects of trial number, target word frequency, the number of words prior to the target and the probability of the target word were observed (in the same direction). The individual-differences model showed no effect of (continuous) age. The discrepancy regarding the age effect between the age group analysis (see Table X) and the individual-differences model in Table XI suggests that multicollinearity was an issue in the latter more complex pupil peak amplitude model (Table XI). As can be seen in Table III, age is correlated with most of the participant-related variables. If correlated variables are fed into the regression analysis simultaneously, variance is inflated resulting in higher standard errors and thus reduced statistical power. We also set up a more parsimonious individual-differences model, leaving out those participant-related variables which were considerably correlated with age ($|r| > 0.60$), i.e., high-frequency hearing loss (PTA[HF]), fluid cognitive processing ability and visual acuity. As expected, age effects reappeared in this model ($|t| = 2.39$, $p < 0.05$), with reduced pupil dilation amplitudes for older participants. The more parsimonious model was similar to the model presented in Table XI in all other respects. To follow up on this we also conducted the individual differences analyses separately for each age group (models not reported in detail here). In line with the overall model (Table XI), none of the age groups showed a speech rate effect. These separate age group models also showed that the effects of trial number, target word predictability and target word position (reported in Table XI) were driven mainly by the younger participants. This may relate to age differences in dynamic range of the task-evoked pupil reaction discussed above.

In sum, no speech rate effects were observed on the dependent variable pupil peak amplitude (research question 1). The data also did not show evidence for age group differences in the effect of speech rate (research question 2). Similarly, none of the included participant-related variables was associated with the size of the speech rate effect on the pupil peak amplitude.

III. DISCUSSION AND CONCLUSION

Speech rate effects in aging have been addressed in numerous studies (e.g., Schmitt and Moore, 1989;

J. Acoust. Soc. Am. **139** (4), April 2016

Xaver Koch and Esther Janse   1633

Gordon-Salant and Fitzgibbons, 1999; Wingfield *et al.*, 1999). Most of these studies have used artificial time compression to systematically vary speech rate. Possibly, the common observation that older adults show stronger speech rate effects than younger adults is (partly) due to signal degradation caused by time compression techniques (Schneider *et al.*, 2005; Gordon *et al.*, 2009; Kusomoto and Vaughan, 2004) or to the fact that many studies have compressed speech to rates that are higher than typically found in natural speech. The present study was set up to investigate speech rate effects on word recognition across the adult life span by using variation in speech rate within and between speakers as found in a corpus of conversational speech. In addition, to address the different accounts that have been put forward for the *age × speech rate* interaction, participant-related variables were collected to study which cognitive, perceptual and linguistic abilities may modulate the size of the speech rate effect on speech recognition. A word recognition task was embedded in a visual-world eye-tracking paradigm, such that multiple dependent variables were obtained at a time (click response times, eye gaze behavior, and pupillometry measures). As expected, increased speech rate made word recognition more challenging as evidenced by longer click response times and delayed eye gaze behavior to the target word. Thus, even though our speech materials were less controlled than artificially speeded lab-recorded sentences, rate variation in our conversational stimuli affected ease of word recognition. Furthermore, age effects were observed on click response times, eye gaze behavior, and on the pupil measures, with slower click responses, slower and decreased pupil dilation responses and slightly delayed gaze behavior for the older adults. However, our main question was whether younger, middle-aged, and older adults differ in the ability to keep up with faster rates of speech. None of the dependent variables under investigation showed any convincing evidence that increased speech rate affected older or middle-aged adults more than younger adults. Relatedly, none of the participant-related measures (e.g., hearing sensitivity or fluid cognitive processing ability) modulated the speech-rate effect on the different indices of word recognition.

Even though we found consistent effects of speech rate on click response times and gaze behavior, these effects were not found in the pupillometry measures. This may be due to our experimental procedure (i.e., the visual-world paradigm). The fact that participants moved their eyes because of the visual search task may have affected peak detection (resulting either in missing values or incorrect peaks). This may have reduced statistical power of our pupillometry analyses. Note that the pupil dilation curves (provided in Fig. 5) suggest that pupil peaks are slightly higher and somewhat delayed for higher speech rates, indexing increased processing effort.

Higher natural speech rates not only present listeners with a higher information rate to keep up with (i.e., more words per minute), but also with word forms that are more

reduced (Ernestus and Warner, 2011) and hence less redundant (cf. Aylett and Turk, 2004). This effect of less clear articulation was also particularly present in the study by Gordon-Salant *et al.* (2014), who found that older and younger adults had more difficulty with naturally produced fast speech than with artificially speeded speech. Whereas younger and older adults showed equal performance for the normal-rate speech in their study, older adults performed more poorly than younger adults both for the time-compressed and naturally fast materials (thus again showing an *age × speech rate* interaction). As argued in the introduction, speech obtained by instructing speakers to read aloud at their ceiling rate (as in Gordon-Salant *et al.*, 2014) may be different from speech varying in tempo as encountered in everyday conversations. In our study, fragments were taken from a corpus of conversational speech in which speakers speak at their habitual rate or deliberately choose to speak at a particular tempo. Possibly, pushing speakers to speak faster than they would normally do (with no communicative intent) may yield more slurring and acoustic reduction than present in our materials. Only more extreme fast and slurred articulation might have affected older adults more adversely than younger adults.

The combined pattern of results thus converges on speech rate effects being similar across age groups for conversational speech fragments. Note that this may be because our older adults had relatively good hearing as they were not eligible for hearing aids. The different accounts of the *age × speech rate* interaction have either emphasized the role of age-related hearing loss or cognitive decline (cognitive slowing in particular). Hearing loss did not affect our dependent variables (in models in which age was also included), nor did it interact with the effect of speech rate. Fluid cognitive processing ability, measuring cognitive slowing, affected click response times and pupil dilation latency in the expected direction, but did not modulate effects of speech rate. Apart from cognitive and hearing abilities, we also expected linguistic abilities to facilitate word recognition. Participants with better vocabulary knowledge were shown to have faster click responses. Thus, speech processing may be facilitated by hearing, cognitive and linguistic abilities, but they were not found to modulate effects of speech rate. Therefore, our findings emphasize that earlier claims about *age × speech rate* interactions mainly obtained with artificial time compression may not generalize to natural speech rate variation as encountered in conversational speech, at least not for an older adult sample with relatively good hearing.

Xaver Koch and Esther Janse

[1]Taking the formula of Lamel *et al.* (1989) to convert words per minute into syllables per second (syll./s) or minute.
[2]*B* here denoting an unstandardized coefficient.
[3]β here denoting a standardized coefficient.
[4]See supplementary material at http://dx.doi.org/10.1121/1.4944032 for transcriptions of all 60 question-answer sequences including item-related characteristics and the orthographic representations used throughout the experiment (target words and distractors).

Allopenna, P. D., Magnuson, J. M., and Tanenhaus, M. K. (**1998**). "Tracking the time course in spoken word recognition using eye movements: Evidence for continuous mapping models," J. Mem. Lang. **38**, 419–439.

Andreassi, J. L. (**2000**). "Pupillometry and behaviour," in *Psychophysiology. Human Behaviour and Physiological Response* (Lawrence Erlbaum, Hillsdale, NJ), pp. 289–307.

Aylett, M., and Turk, A. (**2004**)."The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang. Speech **47**(1), 31–56.

Baayen, R. H., Piepenbrock, R., and Rijn, H. (**1993**). "The CELEX Lexical Database," Linguistic Data Consortium (University of Pennsylvania), electronic resource.

Barr, D. J. (**2008**). "Analyzing 'visual world' eyetracking data using multilevel logistic regression," J. Mem. Lang. **59**, 457–474.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2013**). "lme4: Linear mixed-effect models using Eigen and S4 (R package version 1.0-5)," www.CRAN.R-project.org/package=lme4 (Last viewed September 3, 2014).

Beatty, J. (**1982**). "Task evoked pupillary responses, processing load, and the structure of processing resources," Psychol. Bull. **91**(2), 276–292.

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., and Schneider, B. A. (**2011**). "Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements," J. Speech Lang. Hear. Res. **54**, 243–262.

Birren, J. E., Casperson, R. C., and Botwinick, J. (**1950**). "Age changes in pupil size," J. Gerontol. **5**(3), 216–221.

Brouwer, S., Mitterer, H., and Huettig, F. (**2012**). "Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations," Q. J. Exp. Psychol. **65**(11), 2193–2220.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**(4), 2108–2120.

Carhart, R., and Jerger, J. F. (**1959**). "Preferred method for clinical determination of pure-tone thresholds," J. Speech. Hear. Dis. **24**, 330–345.

Cerella, J. (**1990**). "Aging and information processing rate," in *Handbook of the Psychology of Aging*, edited by J. E. Birren and K. W. Schaie (Academic Press, New York), pp. 201–221.

Chen, D. S., Betz, J., Yaffe, K., Ayonayon, H. N., Kritchevsky, S., Martin, K. R., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Xue, Q., Pratt, S., Simonsick, E. M., and Lin, F. R. (for the Health ABC study) (**2015**). "Association of hearing impairment with declines in physical functioning and the risk of disability in older adults," J. Gerontol. A Biol. Sci. Med. Sci. **70**(5), 654–661.

Coyle, T. R., Pillow, D. R., Snyder, A., and Kochunov, P. (**2011**). "Processing speed mediates the development of general intelligence (*g*) in adolescence," Psychol. Sci. **22**(10), 1265–1269.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A., and Nondahl, D. M. (**1998**). "Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The epidemiology of hearing loss study," Am. J. Epidemiol. **148**(9), 879–886.

Ernestus, M., and Warner, N. (**2011**). "An introduction to reduced pronunciation variants," J. Phonetics **39**, 253–260.

Fitzgibbons, P. J., and Gordon-Salant, S. (**2010**). "Behavioural studies with aging humans: Hearing sensitivity and psychoacoustics," in *The Aging Auditory System. Springer Handbook of Auditory Research* (Springer, Berlin), Vol. 34, pp. 111–134.

Gordon, M. S., Daneman, M., and Schneider, B. A. (**2009**). "Comprehension of speeded discourse by younger and older listeners," Exp. Aging Res. **35**, 277–296.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Lang. Hear. Res. **36**, 1276–1285.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.

Gordon-Salant, S., Yeni-Komishian, G. H., Fitzgibbons, P. J., and Barret, J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.

Gordon-Salant, S., Zion, D. J., and Espy-Wilson, C. (**2014**). "Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners," J. Acoust. Soc. Am. **136**(4), EL268–EL274.

Goy, H., Pelletier, M., Coletta, M., and Pichora-Fuller, M. K. (**2013**). "The effects of semantic context and the type and amount of acoustical distortion on lexical decision by younger and older adults," J. Speech Lang. Hear. Res. **56**(6), 1715–1732.

Greenberg, S. (**1998**). "Speaking in shorthand—A syllable-centric perspective for understanding pronuniciation variation," in *Proceedings of the ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition*, Kekrade, the Netherlands, pp. 47–56.

Haubert, N., and Pichora-Fuller, M. K. (**1999**). "The perception of spoken language by elderly listeners: Contributions of auditory temporal processes," Can. Acoust. **27**, 96–97.

Holladay, J. T. (**1997**)."Proper method for calculating average visual acuity," J. Refract. Surg. **13**, 388–391.

Holube, I., Fredelake, S., Vlamig, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol. **49**(12), 891–903.

Hox, J. J. (**2010**). *Multilevel Analysis: Techniques and Applications* (Routledge, New York).

Humes, L. E., Burk, M. H., Coughling, M. P., Busey, T. A., and Strauser, L. E. (**2007**). "Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effect of presentation rate," J. Speech Lang. Hear. Res. **50**, 283–303.

Janse, E. (**2004**). "Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech," Speech Commun. **42**, 155–173.

Janse, E. (**2009**). "Processing of fast speech by elderly listeners," J. Acoust. Soc. Am. **125**(4), 2361–2373.

Jiang, Z. (**2010**). www.mathworks.com/matlabcentral/fileexchange/6271-spec-file-reader/content/specreader/smooth.m (Last viewed February 10, 2016).

Just, M. A., and Carpenter, P. A. (**1980**). "A theory of reading: From eye fixation to comprehension," Psychol. Rev. **87**, 329–354.

Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**(5), 1208–1221.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**(1), 362–378.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., and Eckert, M. A. (**2013**). "Pupil size varies with word listening and response selection difficulty in oder adults with hearing loss," Psychophysiology **50**(1), 23–34.

Kusomoto, A., and Vaughan, N. (**2004**). "Comparison of acoustic features of time-compressed and natural speech," J. Acoust. Soc. Am. **116**(4), 2600.

Lambrev, P. (**2015**). www.mathworks.com/matlabcentral/fileexchange/32828-spectr-o-matic/content/peakdet.m (Last viewed February 10, 2016).

Lamel, L. F., Kassel, R. H., and Seneff, S. (**1989**). "Speech database development: Design and analysis of the acoustic-phonetic corpus," in *Speech I/O Assessment and Speech Databases (ISCA)*, Noordwijkerhout, the Netherlands, pp. 2161–2170.

Luteijn, F., and van der Ploeg, F. A. E. (**1983**). *Handleiding Groninger Intelligentietest (Manual Groningen Intelligence Test)* (Swets and Zeitlinger, Lisse, the Netherlands).

McQueen, J. M., and Viebahn, M. C. (**2007**). "Tracking recognition of spoken words by tracking looks to printed words," Q. J. Exp. Psychol. **60**(5), 661–671.

Mirman, D. (**2014**). *Growth Curve Analysis and Visualization in R* (CRC Press, Boca Raton, FL).

Noble, W., and Gatehouse, S. (**2004**). "Interaural asymmetry of hearing loss, speech, spatial and quality of hearing scale (SSQ) disabilities, and handicap," Int. J. Audiol. **43**, 100–114.

Oostdijk, N. (**2000**). "The Spoken Dutch Corpus. Overview and first evaluation," in *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, Paris, France, pp. 887–893.

Park, J., Carp, J., Hebrank, A., Park, D. C., and Polk, T. A. (**2010**). "Neural specificity predicts fluid processing ability in older adults," J. Neurosci. **30**(27), 9253–9259.

Pichora-Fuller, M. K., Schneider, B., Benson, N., Hamstra, S., and Storzer, E. (**2006**). "Effect of age on gap detection in speech and non-speech stimuli varying in marker duration and spectral symmetry," J. Acoust. Soc. Am. **119**(2), 1143–1155.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., and Agular, M. (**2008**). "The pupil dilation response to visual detection," in *Proceeding of SPIE-IS&T Human Vision and Electronic Imaging XIII*, edited by B. E. Rogowitz and T.N. Pappas, Vol. 4806.

Quené, H. (**2008**). "Multilevel modeling of between-speaker and within speaker variation in spontaneous speech tempo," J. Acoust. Soc. Am. **123**(2), 1104–1113.

Quené, H. (**2014**). (private communication).

Raphael, L. J., and Dorman, M. F. (**1980**). "Silence as a cue to the perception of syllable-initial and syllable-final stop consonants," J. Phonetics **8**, 269–275.

Raven, J., Raven, J. C., and Court, J. H. (**2003**). *The Manual for the Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview* (Harcourt Assessment, San Antonio).

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (**2013**). "The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances," Front. Syst. Neurosci. **7**, 1–17.

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (**2008**). "Cognition counts: A working memory system for ease of language understanding (ELU)," Int. J. Audiol. **47**(Suppl. 2), S99–S105.

Salthouse, T. A. (**1985**). "Speed of behaviour and its implications for cognition," in *Handbook of the Psychology of Aging*., edited by J. Birren, K. Schaie, and K. Warner (Van Nostrand Reinhold, New York), pp. 400–426.

Salthouse, T. A. (**1996**). "The processing-speed theory of adult age differences in cognition," Psychol. Rev. **103**(3), 403–428.

Salthouse, T. A. (**2000**). "Aging and measures of processing speed," Biol. Psychol. **54**, 35–54.

Schmidtke, J. (**2014**). "Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry," Front. Psychol. **5**, 1–16.

Schmitt, J. F., and Moore, J. R. (**1989**). "Natural alteration of speaking rate: The effect on passage comprehension by listeners over 75 year of age," J. Speech Hear. Res. **32**, 445–450.

Schneider, B. A., Daneman, M., and Murphy, D. R. (**2005**). "Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?," Psychol. Aging **20**(2), 261–271.

Small, J. A., Andersen, E. S., and Kempler, D. (**1997**). "Effects of working memory on understanding rate-altered speech," Aging Neuropsychol. C **4**(2), 126–139.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K., and Levinson, S. C. (**2011**). "Universals and cultural variation in turn-taking in conversation," Proc. Natl. Acad. Sci. U.S.A. **106**(26), 10587–10592.

van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., and Schmidt, H. G. (**2004**). "Memory load and the cognitive pupillary response in aging," Psychophysiology **41**, 167–174.

Vaughan, N., Storzbach, D., and Furukawa, I. (**2006**). "Sequencing versus nonsequencing working memory in understanding of rapid speech by older listeners," J. Am. Acad. Audiol. **17**, 506–518.

Wechsler, D. (**2004**). *Wechsler Adult Intelligence Test*, 3rd ed., Dutch version (Harcourt Test, Amsterdam).

Wilhelm, O., and Schulze, R. (**2002**). "The relation of speeded and unspeeded reasoning with mental speed," Intelligence **30**, 537–554.

Wingfield, A. (**1996**). "Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory," J. Am. Acad. Audiol. **7**(3), 175–182.

Wingfield, A., Peelle, J. E., and Grossman, M. (**2003**). "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," Aging Neuropsychol. Cogn. **10**(4), 310–322.

Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (**1999**). "Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech," Psychol. Aging **14**(3), 380–389.

Zekveld, A. A., Festen, J. M., and Kramer, S. E. (**2013**). "Task difficulty differentially affects two measures of processing load: The pupil response during sentence processing and delayed cued recall of the sentences," J. Speech Lang. Hear. Res. **56**, 1156–1165.

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2010**). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," Ear Hear. **31**(4), 480–490.

Zipf, G. K. (**1949**). *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Oxford).

Zipf, G. K. (**1965**). *The Psychobiology of Language. An Introduction to Dynamic Philology* (MIT Press, Cambridge, MA).