Cognitive load makes speech sound fast, but does not modulate acoustic context effects

Hans Rutger Bosker [a,b,*], Eva Reinisch [c], Matthias J. Sjerps [b,d]

[a] Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[c] Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany
[d] Department of Linguistics, University of California, Berkeley, CA, USA

ARTICLE INFO

ABSTRACT

In natural situations, speech perception often takes place during the concurrent execution of other cognitive tasks, such as listening while viewing a visual scene. The execution of a dual task typically has detrimental effects on concurrent speech perception, but how exactly cognitive load disrupts speech encoding is still unclear. The detrimental effect on speech representations may consist of either a general reduction in the robustness of processing of the speech signal ('noisy encoding'), or, alternatively it may specifically influence the temporal sampling of the sensory input, with listeners missing temporal pulses, thus underestimating segmental durations ('shrinking of time'). The present study investigated whether and how spectral and temporal cues in a precursor sentence that has been processed under high vs. low cognitive load influence the perception of a subsequent target word. If cognitive load effects are implemented through 'noisy encoding', increasing cognitive load during the precursor should attenuate the encoding of both its temporal and spectral cues, and hence reduce the contextual effect that these cues can have on subsequent target sound perception. However, if cognitive load effects are expressed as 'shrinking of time', context effects should not be modulated by load, but a main effect would be expected on the perceived duration of the speech signal. Results from two experiments indicate that increasing cognitive load (manipulated through a secondary visual search task) did not modulate temporal (Experiment 1) or spectral context effects (Experiment 2). However, a consistent main effect of cognitive load was found: increasing cognitive load during the precursor induced a perceptual increase in its perceived speech rate, biasing the perception of a following target word towards longer durations. This finding suggests that cognitive load effects in speech perception are implemented via 'shrinking of time', in line with a temporal sampling framework. In addition, we argue that our results align with a model in which early (spectral and temporal) normalization is unaffected by attention but later adjustments may be attention-dependent.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Speech perception is most commonly studied under ideal listening conditions that allow participants to dedicate their full attention to the listening task at hand. However, natural conversations typically take place in a world

* Corresponding author at: Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands.
  E-mail address: HansRutger.Bosker@mpi.nl (H.R. Bosker).

where multiple cognitive tasks compete for limited central processing resources. Listening to speech when cognitive resources are distributed across multiple tasks is referred to as listening under cognitive load. Such cognitive load can be imposed by any additional attentional or mnemonic process, and specifically excludes any effect on speech perception that arises from an energetic distortion of the signal. Increasing cognitive load typically has detrimental effects on speech perception (e.g., worse phoneme monitoring accuracy, Wurm & Samuel, 1997; worse word segmentation; Fernandes, Kolinsky, & Ventura, 2010), but the underlying mechanism responsible for adverse influences of cognitive load is debated. This study investigates two potential mechanisms proposed in the literature and provides empirical support that cognitive load influences the temporal computation of the sensory input.

Increases in cognitive load due to a dual task (e.g., difficult visual search) are known to result in increased reliance on lexical relative to acoustic information in phonetic categorization. Mattys and Wiget (2011) presented listeners with a *giss-kiss* continuum and observed that identification responses for the word-initial consonant showed a stronger lexical bias (i.e., more /k/ responses, "Ganong effect"; Ganong, 1980) under increased cognitive load. They argue that cognitive load has detrimental effects on speech sound representations in early stages of sensory input analysis, causing listeners to rely more strongly on information about the unaffected lexical representations. That is, sub-lexical (phonetic) encoding appears to be disrupted under cognitive load (Mattys, Barden, & Samuel, 2013; Mattys & Palmer, 2015).

At least two perceptual mechanisms have been suggested to explain how increases in cognitive load induce impoverished phonetic encoding. One mechanism (henceforth, the 'noisy encoding' mechanism) suggests that cognitive load negatively affects speech perception because of a decrease in the perceptual 'signal-to-noise' ratio; that is, a reduction in the strength of the pre-lexical representations compared to the level of the background 'system noise'. To exemplify, within a Signal Detection framework, the role of attention in perception may be conceived of as modulating the signal-to-noise ratio in the perceptual system, giving contrastive cues priority over non-contrastive cues (e.g., Gordon, Eberhardt, & Rueckl, 1993). This less favorable signal-to-noise ratio could come about as a result of a decrease in the strength of processing of the speech cues and/or by a failure to suppress/filter-out system noise, resulting in the masking of relevant speech cues. Since this framework assumes random system noise, cognitive load would be expected to have a general detrimental effect on the encoding of any kind of phonetic cue (i.e., affecting the perception of both spectral and temporal characteristics of the signal).

Another mechanism (henceforth, the 'shrinking of time' mechanism) that may underlie detrimental cognitive load effects involves sensory time perception. Arguing from a domain-general timer hypothesis (e.g., Coull, Vidal, Nazarian, & Macar, 2004; Macar, Grondin, & Casini, 1994), estimates of the duration of sensory input are based on the registration of temporal pulses. Increasing cognitive load may decrease the sampling rate of input processing,

causing listeners to miss temporal pulses, leading to a loss of sensory information. This mechanism is supported by findings that duration judgments under cognitive load result in systematic underestimation of time as cognitive load increases (Block, Hancock, & Zakay, 2010: a meta-analysis of 117 experiments). That is, the more one's attentional resources are taxed, the faster time seems to pass, and duration estimates of any sensory input received during that time are shortened.

Importantly, this 'shrinking of time' has been shown to affect the perception of speech sounds (Casini, Burle, & Nguyen, 2009). Casini et al. presented French participants with a /ʃ/-/ʒ/ voicing continuum in French, a distinction that is partly cued by the duration of the preceding vowel. When their participants performed phonetic identification of this contrast under cognitive load, they were biased towards perceiving /ʃ/ (cued by a shorter vowel). This finding thus aligns with the notion that cognitive load caused an underestimation of the perceived vowel duration.

Note that these two mechanisms ('noisy encoding' vs. 'shrinking of time') are by no means mutually exclusive, and both concepts are instructive to our understanding of the influence of cognitive load on speech perception. However, to our knowledge, no study has directly compared the two in a single experimental paradigm. Both mechanisms suggest that effects of cognitive load operate at an early locus in perception, affecting the initial perceptual encoding of low-level phonetic cues. However, the two mechanisms differ with respect to their specificity. The 'noisy encoding' mechanism predicts that an increase in cognitive load leads to general disruptions in phonetic encoding, inducing weaker representation of any phonetic cue in the speech signal (i.e., both spectral and temporal speech cues, hence leading to an increased reliance on other cues; Mattys & Wiget, 2011). The 'shrinking of time' mechanism is somewhat more specific in proposing that sparser temporal sampling underlies cognitive load effects. This predicts that only temporal encoding of speech should be disrupted (i.e., underestimation of segmental durations) while the perception of spectral cues remains unaffected.

In order to investigate the involvement of these two mechanisms in speech perception, the present study investigated how cognitive load affects the influence that spectral and temporal cues in a *precursor* sentence have on the perception of a subsequent target word. The acoustic context in which speech sounds occur has long been known to affect their perception. For example, the spectral content of a sentence contrastively influences the perception of a subsequent target word (e.g., Ladefoged & Broadbent, 1957). Ladefoged and Broadbent demonstrated that the perception of an /ɛ/-/ɪ/ continuum can be shifted towards /ɪ/ (lower first formant, F1) by presenting it in a sentence with relatively high F1. A similar contrastive influence has been reported for the temporal properties of acoustic context (e.g., Pickett & Decker, 1960), with segmental durations being perceived as longer when the surrounding speech rate is increased. These two types of acoustic context effects are known as *spectral normalization* and *rate normalization*.

Acoustic context effects have been suggested to be largely caused by general auditory processes that occur at

early stages in perception. This is supported by findings that (1) non-speech, such as pure tones or sine wave speech, can also trigger acoustic context effects (Bosker, 2016b; Diehl & Walsh, 1989; Gordon, 1988; Huang & Holt, 2009; Laing, Liu, Lotto, & Holt, 2012; Sjerps, Mitterer, & McQueen, 2011; Stilp, Alexander, Kiefte, & Kluender, 2010; Wade & Holt, 2005); (2) non-human auditory perception exhibits qualitatively similar context effects (Dent, Brittan-Powell, Dooling, & Pierce, 1997; Lotto, Kluender, & Holt, 1997; Sinnott, Brown, & Borneman, 1998; Welch, Sawusch, & Dent, 2009); and (3) acoustic context effects occur very rapidly (Reinisch & Sjerps, 2013; Toscano & McMurray, 2015) operating prior to other perceptual processes, such as lexically guided perceptual learning (Sjerps & Reinisch, 2015) and stream segregation (Bosker, 2016a; Newman & Sawusch, 2009; cf. Reinisch, 2016b). Since an important part of acoustic context effects operates at an early locus in perception (similar to the aforementioned effects of cognitive load), cognitive load may be expected to influence the way acoustic context affects speech perception.

Crucially, the two mechanisms suggested to underlie the disruptive effects of cognitive load on speech perception ('noisy encoding' and 'shrinking of time') make different -orthogonal- predictions about how cognitive load should influence these acoustic context effects. Consider a situation in which a listener is presented with a precursor sentence under increased cognitive load (e.g., while performing a concurrent visual search task). According to the 'noisy encoding' mechanism, the signal-to-noise ratio in the activation of acoustic/phonetic representations would be diminished, resulting in less robust representations. As a consequence, both the spectral and temporal characteristics of the precursor would be expected to exert less of an influence on subsequent target perception. Thus one would expect both rate normalization effects and spectral normalization effects to be reduced under increased cognitive load.

This hypothesis aligns with findings that diverting the focus of auditory attention away from a particular speech stream (a) decreases the robustness of speech stream representations at cortical areas encoding phonetic information (Mesgarani & Chang, 2012) and (b) reduces cortical tracking of the slow amplitude modulations in speech (Golumbic et al., 2013; Kerlin, Shahin, & Miller, 2010) which has been suggested to play a central role in rate normalization (Bosker, 2016b; Peelle & Davis, 2012). Also, impoverishing the speech signal while the availability of contextual cues is maintained, reduces (spectral and rate) normalization effects (Gordon, 1988; Sjerps et al., 2011).

In contrast, a 'shrinking of time' mechanism predicts that cognitive load induces sparser temporal sampling of the spoken input, with listeners failing to register temporal pulses. This, in turn, would induce a perceived increase in the perceptual speech rate of the precursor sentence. This increase in perceived speech rate could elicit an independent rate normalization effect. That is, segmental durations following the precursor should be perceived as longer because the preceding speech was perceived to have a perceptually faster speech rate – similar to how actually pro-

duced fast speech biases perception of following speech towards longer segments (relative to slow speech).

Crucially, a 'shrinking of time' mechanism does *not* predict a modulation of spectral normalization or rate normalization effects. Instead, it suggests that the perceptual increase in speech rate (induced by sparser temporal sampling as a result of the cognitive load manipulation) operates over and above any effects of the actual spectral and temporal cues in the precursor. That is, all sentences may be perceived as faster under higher cognitive load (regardless of any spectral or temporal acoustic manipulations), biasing perception of subsequent target segments toward longer durations.

To test whether and how cognitive load affects the strength of contextual influences, two experiments were designed. In Experiment 1, we investigated effects of cognitive load on temporal context effects (rate normalization). In Experiment 2, we investigated effects of cognitive load on spectral context effects (spectral normalization). In both experiments, participants were presented with speech stimuli that consisted of manipulated precursor sentences that ended in target words containing vowels ambiguous between Dutch /ɑ/ and /aː/. This vowel contrast is cued by both spectral (lower formant values for /ɑ/, higher formant values for /aː/) and temporal cues (shorter duration for /ɑ/, longer duration for /aː/; Escudero, Benders, & Lipski, 2009; Gerrits, 2001; Nooteboom & Cohen, 1984; van Heuven, Van Houten, & De Vries, 1986) to a similar degree. As such, this vowel contrast is an ideal test contrast for our purposes: it is susceptible to both temporal and spectral context effects (faster speech rates and lower second-formant values in the acoustic context shift perception towards /aː/; Reinisch & Sjerps, 2013).

During the presentation of the manipulated precursor sentences (*not* during presentation of the target contrast), participants performed a visual search task that was either difficult ("high load") or easy ("low load"). That is, the current experiment was designed to leave the processing of the actual target sound unaffected. Instead we manipulated the encoding of a potential secondary cue; the information present in a stretch of preceding context.

The cognitive load manipulation allowed us to test for (1) a potential modulation of acoustic context effects, and (2) potential overall shifts in perception related to the proposed 'shrinking of time'. Specifically, if an increase in cognitive load disrupts the low-level encoding of speech through a lower signal-to-noise ratio in the activation of phonetic representations ('noise encoding'), we would expect to find both a reduction in the rate normalization effect in Experiment 1, and a reduction in the spectral normalization effect in Experiment 2. Alternatively, if an increase in cognitive load disrupts speech encoding by missing temporal pulses ('shrinking time'), we would expect to find a main effect of cognitive load in both experiments. That is, if cognitive load induced a perceptual increase in the perceived speech rate of the precursor sentence independent of any rate or spectral manipulation between sentences, we would predict to find a higher proportion of /aː/ (long vowel) responses in the high load condition than in the low load condition. This effect may be

independent of or in addition to any contextual effect of the actual spectral and temporal characteristics of the precursors.

Experiment 1: rate normalization under cognitive load

Method

Participants

Native Dutch participants (N = 32) with normal hearing were recruited from the Max Planck Institute's participant pool. They gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). Data from 4 participants were excluded for reasons of fatigue or non-compliance, leaving data from 28 participants (5 males, M_{age} = 25) for analysis.

Design and materials

The dual-task design of the experiment was modeled after Mattys and Wiget (2011). Stimuli used for the visual search task consisted of object grids, containing an equal number of randomly positioned red diamonds, red triangles, black squares, and black upside-down triangles. A grid of 4 rows and 4 columns made up the low load condition and a grid of 13 rows and 13 columns made up the high load condition (see examples in Fig. 1). In half of the trials, one randomly selected object in the grid was replaced by an oddball object. The oddball was always a black diamond.

The auditory stimuli in the experiment were adopted from Reinisch and Sjerps (2013) to which the reader is referred for details beyond what will be described here. Each stimulus consisted of the same semantically unbiasing sentence followed by a target word: "Klik nu een keer op het woord [target]"; *This time click on the word [target].* The sentence, with an original duration of 1220 ms (5.74 syllables per second), was linearly compressed to 66% of its original duration for a fast version (793 ms; 8.83 syllables per second) and linearly expanded to 133% for a slow

version (1648 ms; 4.25 syllables per second). The rate change was implemented using PSOLA in Praat (Boersma & Weenink, 2016). The sentence did not contain any /ɑ/ or /aː/ vowels since these made up the critical contrast for the target words. Target pairs were "zak" - "zaak", *bag - case*; "gas" - "gaas", *gas - mesh*; "macht" - "maagd", *power - virgin* (for details on how these target pairs were chosen see Reinisch & Sjerps, 2013).

For each target pair, a separate vowel duration continuum was created consisting of three different vowel duration steps (short, mid, long), all falling within the natural range of our particular speaker. Vowels were shortened or lengthened manually by removing or duplicating individual pitch periods throughout the vowels. The duration continuum spanned 40 ms, from 120 ms to 160 ms, in steps of approximately 20 ms as permitted by the duration of the individually removed periods. The spectral characteristics of the three vowel steps (F1 and F2) were kept constant at ambiguous values as established in the previous study. Categorization data from Reinisch and Sjerps (2013) showed that the duration continua for each target pair sampled similar points on a perceptual /ɑ/-/aː/ categorization curve: token 1 (short duration) received an average percentage /aː/ responses of 8%, token 2 (medium duration) received an average percentage /aː/ responses of 38%, and token 3 (long duration) received an average percentage /aː/ responses of 76%. Target words were combined with the fast and slow sentences yielding a total of 18 unique stimuli.

Procedure

Stimulus presentation was controlled by Presentation software (v16.5; Neurobehavioral Systems, Albany, CA, USA). The two cognitive load conditions were blocked, with block order counter-balanced across participants. Each block consisted of 216 trials, with each of the 18 unique stimuli presented 12 times within one block. In half of the trials, there was an oddball present in the visual grid. Speech stimuli within a block were presented in a fixed random order to half of the participants, with the reversed order presented to the other half. Between the two blocks participants were allowed to take a short break.

Fig. 2 illustrates the time-course of one trial. Each trial started with a fixation cross appearing in the middle of the screen. After 500 ms, an object grid was visually presented (small grid in the low load block, large grid in the high load block). Participants were allowed a preview time of 250 ms, after which the spoken sentence was presented. At precursor offset, the object grid was replaced by a blank screen. This meant that participants performed the visual search task *only during precursor presentation*, not during the perception of the target word. Because the precursor's duration was dependent on the specific rate condition of the trial, the search time given to participants varied with the precursor's speech rate. That is, participants had a shorter search time (793 ms) for fast precursors than for slow precursors (1648 ms). Since we were interested in the overall difference between the high and the low load conditions, however, this was considered non-problematic (i.e., within both short and long precursor conditions, high load always required more searching than low
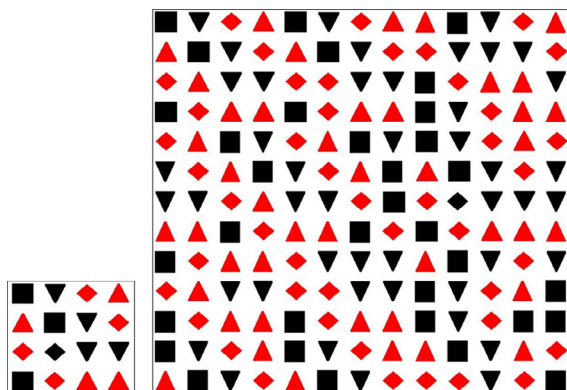

Fig. 1. Examples of the object grids used for the visual search task. The left panel shows a 4 × 4 grid used for the low load condition and the right panel a 13 × 13 grid used for the high load condition. Both grids shown here are examples of trials with the oddball object (the black diamond) being present.
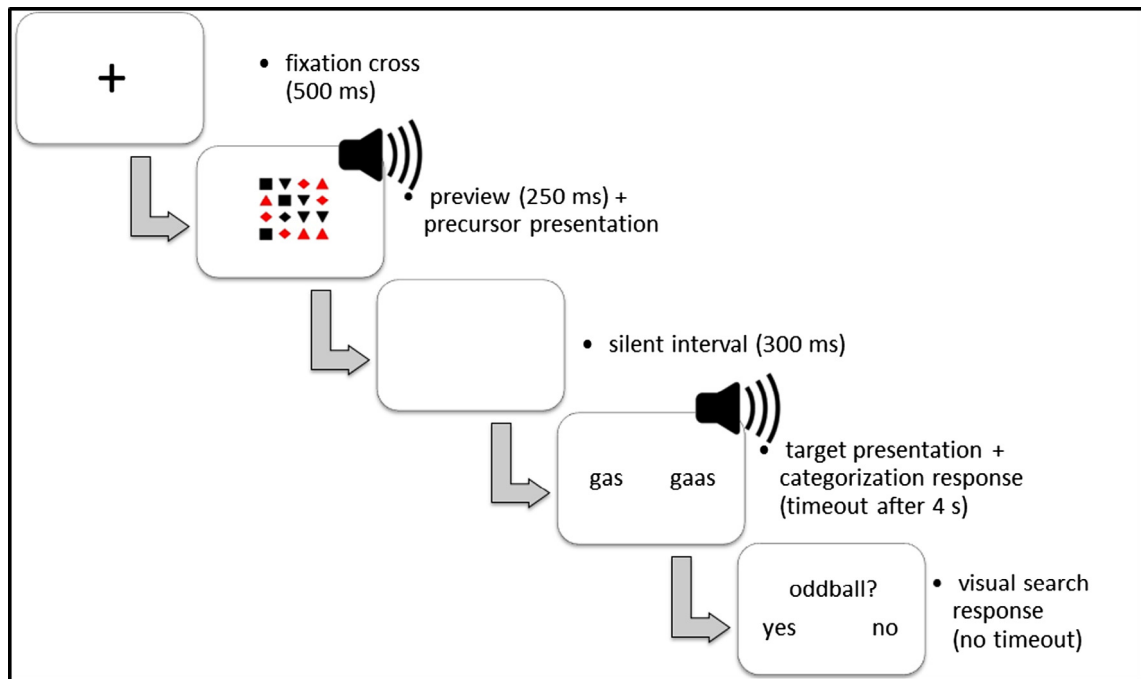
Fig. 2. Schematic representation of trial design. A trial started with a fixation cross, presented for 500 ms, followed by the visual presentation of an object grid (4 × 4 grid in the low load condition; 13 × 13 grid in the high load condition). After a preview of 250 ms, the spoken precursor was presented. At precursor offset, the object grid was replaced by a blank screen for 300 ms. Then, the target word was presented concurrently with two visually presented response options on either side of the screen. After participants' categorization response, or after a timeout of 4 s, participants were asked whether or not they had seen the oddball object, the black diamond, in the object grid.

load). After a silent interval of 300 ms, the target word was presented together with two visually presented response options (the two possible words, e.g., *gas* vs. *gaas*, etc.) on either side of the screen (position counter-balanced across participants). Participants pressed the "1" key on the computer keyboard for the left word and "0" for the right word. In case participants did not respond within 4 s after target onset, a missing response was recorded. After participants had logged their response (or after timeout), participants were asked whether or not they had seen the oddball object (the black diamond) in the object grid. They pressed the "J" key for yes (Dutch: *ja*) and "N" for no (Dutch: *nee*). Participants could only proceed to the next trial after pressing one of these two keys (i.e., no timeout; no missing responses).

Results

Trials with missing categorization responses (*n* = 22; <1%) were excluded from analyses. As presented in Table 1,

Table 1
Average performance (mean (*SD*); in percentages) on the visual search task in Experiment 1, split by the two load conditions and the two rate conditions. Chance level is 50%.

	Fast precursor	Slow precursor
High load condition	63 (48)	72 (45)
Low load condition	95 (21)	96 (20)

participants' visual search accuracy was high in the low load condition (close to ceiling) but considerably lower in the high load condition, though well above chance (chance performance is 50%). This suggests that our cognitive load manipulation was successful. In fact, the accuracy scores in
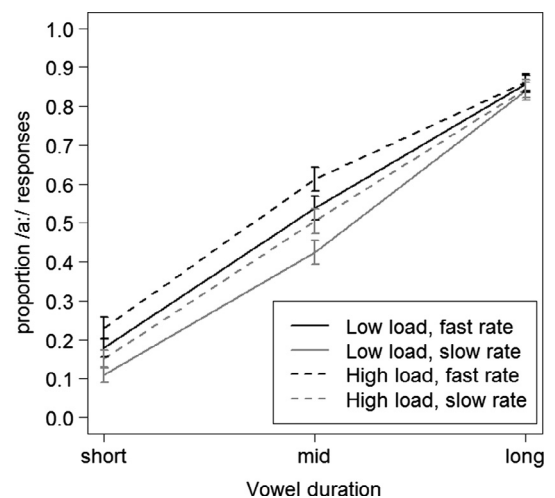

Fig. 3. Average categorization data (in proportion /aː/ responses) for Experiment 1 for different durations from the vowel continua, split by the two precursor conditions, and the two load conditions. Error bars enclose 1.96 × *SE* on either side, 95% CIs.

the high load condition are lower than those reported by Mattys and Wiget (2011), which were close to 80%.

Categorization data, calculated as the proportion of /a:/ responses, are presented in Fig. 3. As expected, a decrease in target vowel duration led listeners to report fewer /a:/ responses, across both rate and load conditions. The difference between the black and grey lines indicates an influence of the precursor's speech rate, with faster speech rates (black lines) biasing perception towards /a:/. Importantly, the difference between the solid and dashed lines seems to show an influence of cognitive load, with higher load also biasing perception towards /a:/. The effect of the precursor's speech rate seems to be similar in both load conditions.

We quantified these effects using a Generalized Linear Mixed Model (GLMM; Quené & Van den Bergh, 2008) with a logistic linking function as implemented in the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) in R (R Development Core Team, 2012). The dependent variable was response /a:/ (coded as 1) or /ɑ/ (coded 0). Fixed effects were Vowel Duration (continuous predictor, scaled around the mean), Precursor Rate (categorical predictor, with slow speech rate coded as −0.5 and fast speech rate as +0.5), Load Condition (categorical predictor, with low load coded as −0.5 and high load coded as +0.5), and all of their interactions. The use of deviation coding of two-level categorical factors (i.e., coded with +0.5 and −0.5) allows us to test main effects of these predictors, since with this coding the grand mean is mapped onto the intercept. Participant was entered as a random factor with by-participant random slopes for all fixed effects and all interactions (Barr, Levy, Scheepers, & Tily, 2013).

This model revealed a significant effect of Vowel Duration ($\beta = 4.489$, $z = 15.099$, $p < .001$), with longer vowel durations increasing the proportion of /a:/ responses. The effect of Precursor Rate ($\beta = 0.549$, $z = 6.692$, $p < .001$) indicated that the faster the precursor's speech rate, the higher the proportion of /a:/ responses. Critically, there was an effect of Load Condition ($\beta = 0.437$, $z = 3.058$, $p = .002$), demonstrating that the higher the cognitive load, the higher the proportion of /a:/ responses. No interaction between Load Condition and Precursor Rate was found ($p > .7$). We found an interaction between Vowel Duration and Precursor Rate ($\beta = −0.668$, $z = −4.307$, $p < .001$) indicating a reduced effect of Precursor Rate for longer vowel tokens. This was most likely induced by a 'ceiling effect' for vowels with a longer duration. That is, at 90% /a:/ responses the vowel could be considered as acoustically unambiguous leaving little room for context effects.

Discussion

Experiment 1 was designed to test to what extent cognitive load influences temporal context effects in speech perception (i.e., rate normalization). Two possible hypotheses were tested. The first hypothesis was that cognitive load would modulate rate normalization (similar to studies testing effects on lexical processing, e.g., Mattys & Wiget, 2011). The reasoning was that if cognitive load negatively affected the signal-to-noise ratio in phonetic encoding, and thus the robustness of perceptual representations

(i.e., the 'noisy encoding' mechanism), one would expect impoverished encoding of the temporal properties of the sentence precursors. As such, the effect of surrounding speech rate should have been attenuated by higher cognitive load. However, no interaction between effects of speech rate and load condition was found.

The second hypothesis was that an increase in cognitive load during the sentence precursors would lead to an increase in the perceived speech rate of those sentences. This hypothesis was confirmed: a main effect of cognitive load was observed, showing a consistently higher proportion of /a:/ responses in the high than in the low load condition. Thus, the perceived rate of the precursor sentences is perceptually increased due to the 'shrinking of time' under cognitive load, in line with earlier findings in the time perception literature (Block et al., 2010; Casini et al., 2009). This faster perceptual rate in the precursor, in turn, exerted a contextual influence on the subsequent target word, biasing perception towards /a:/.

Thus, while the findings did not support predictions according to a 'noisy encoding' mechanism of cognitive load effects, they did align with the idea of a perceptual 'shrinking of time'. However, Experiment 1 only focused on contextual influences of temporal cues (i.e., speaking rate). Maybe the previously reported adverse influence of cognitive load on speech perception resulted from a decrease in robustness of spectral processing only. To test for this possibility, a second experiment was designed targeting another form of acoustic context effect: spectral normalization. Recall that the /ɑ/-/a:/ contrast in Dutch is cued equally by both temporal (short vs. long) and spectral properties (lower vs. higher formant values, respectively). As such, it is also susceptible to both rate and spectral context effects (with lower formant values in the context biasing perception towards /a:/; Reinisch & Sjerps, 2013).

Experiment 2: spectral normalization under cognitive load

As in Experiment 1, this experiment allowed us to test effects of cognitive load in speech perception. If cognitive load disrupts the phonetic encoding of the spectral content in the precursor sentences, this would predict an attenuation of spectral normalization effects under cognitive load. In addition, serving as a replication of Experiment 1, if cognitive load induces the perception of a faster speech rate in the precursors, then a general bias towards /a:/ would also be predicted in Experiment 2, operating independently of any contextual influences of the spectral characteristics of the precursor. Note that the target vowel contrast involves the same vowel contrast (/ɑ/ and /a:/) as in Experiment 1. Although the precursors in this experiment were only manipulated in their spectral domain, this does not prevent listeners from interpreting the contrast in relation to temporal cues as well, here in relation to changes in perceived duration due to shrinking of time.

Finally, if the effects of cognitive load are to align with the idea of a perceptual 'shrinking of time', then combined analyses of the data from Experiments 1 and 2 should reveal similar effect sizes of the cognitive load manipula-

tions across the two experiments. Moreover, combined analyses allow us to test the correlation between the extent to which individual participants suffered from the load manipulation in their visual search accuracy, and individual participants' load effect in target word categorization.

Method

Participants

Native Dutch participants ($N = 30$) with normal hearing were recruited from the Max Planck Institute's participant pool. None of them had participated in Experiment 1. They gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). Data from 4 participants were excluded for reasons of fatigue or non-compliance, leaving data from 26 participants (10 males, $M_{age} = 25$) for analysis.

Design and materials

The design of Experiment 2 was identical to that of Experiment 1, except that the spectral characteristics of the spoken stimuli were manipulated. Materials were again adopted from Reinisch and Sjerps (2013). For details about the stimuli we again refer to this paper. The main characteristics are as follows: Instead of presenting precursors with varying speech rate, now two precursors were selected with vowels that had their F2 increased or decreased by 200 Hz. The duration of the two precursors was constant at 1220 ms, lying in between the fast and slow versions from Experiment 1.

Instead of presenting target vowels with varying durations, this time spectral /ɑ/-/aː/ continua of each target pair were used while keeping vowel duration constant (at an ambiguous value of 140 ms as established in the previous study). We varied F2 instead of duration, so as to encourage participants to rely on spectral properties as much as possible (although, as mentioned above, this does not prevent listeners of relying on perceptual differences in duration as well; cf. Bosker, 2016a). Vowel token 2 from Experiment 1 was taken as reference, and two other vowel tokens with a lower F2 and a higher F2 were selected. Different F2 values were selected for the different target pairs to match the different target pairs on perceptual vowel categorization. The F2 continuum of the vowels in "gas" - "gaas" and "macht" - "maagd" included the steps: token 1, $F2 = 1150$ Hz; token 2, $F2 = 1225$ Hz; and token 3, $F2 = 1300$ Hz. The F2 continuum of the vowel in "zak" - "zaak" included the steps: token 1, $F2 = 1225$ Hz; token 2, $F2 = 1300$ Hz; and token 3, $F2 = 1375$ Hz. As described in detail in Reinisch and Sjerps (2013), these three F2 continua are perceived similarly when presented in isolation: token 1, average percentage /aː/ categorization = 18%; token 2, average percentage /aː/ categorization: 38%; token 3, average percentage /aː/ categorization: 83%. Again, all target words were combined with both types of precursors including a gap of 300 ms.

Table 2

Average performance (mean (SD); in percentages) on the visual search task in Experiment 2, split by the two load conditions and the two F2 conditions. Chance level is 50%.

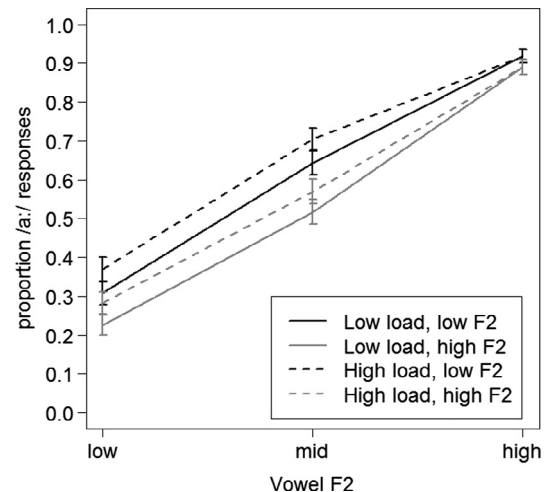|	High F2	Low F2
High load condition	67 (47)	66 (47)
Low load condition	93 (25)	93 (25)

Fig. 4. Average categorization data (in proportion /aː/ responses) for Experiment 2 for different F2s from the vowel continua, split by the two precursor conditions, and the two load conditions. Error bars enclose $1.96 \times SE$ on either side, 95% CIs.

Procedure

The procedure used for Experiment 2 was identical to that of Experiment 1. The visual search task was performed up to precursor offset, not during target presentation. Note that, in contrast to Experiment 1, the precursors used in Experiment 2 did not vary in their duration, but only in their spectral characteristics. As such, the time given to participants to perform the visual search task was constant for all trials of Experiment 2, independent of the precursor condition.

Results

Trials with missing categorization responses ($n = 7$; <1%) were excluded from analyses. As presented in Table 2, participants' visual search accuracy was comparable to that in Experiment 1: Accuracy was high in the low load condition (close to ceiling) and considerably lower in the high load condition, although again well above chance.

Categorization data from Experiment 2, calculated as the proportion of /aː/ responses, are presented in Fig. 4. Differences across continuum steps suggest that the spectral characteristics of the target vowels influenced listeners' perception, similar to the vowel duration continua in Experiment 1. The difference between the black and grey lines suggests that the spectral characteristics of the precursor also biased perception, with precursors with a lower F2 (black lines) biasing perception towards /aː/. Moreover,

the difference between solid and dashed lines suggests an effect of cognitive load that seems comparable across the spectral precursor manipulations, with higher load biasing perception towards /aː/.

A Generalized Linear Mixed Model (GLMM) with a logistic linking function, similar to the one described for Experiment 1, was used to test the binomial responses from Experiment 2 (/aː/ coded as 1). Fixed effects of Vowel F2 (continuous predictor, scaled around the mean), Precursor F2 (categorical predictor, with low F2 precursors coded as −0.5 and high F2 precursors as +0.5), Load Condition (categorical predictor, with low load coded as −0.5 and high load coded as +0.5), and all their interactions were included. Participant was entered as a random factor with by-participant random slopes for all fixed effects and all interactions (Barr et al., 2013).

This model revealed significant effects of Vowel F2 ($\beta = 4.222$, $z = 10.524$, $p < .001$), with higher F2s increasing the proportion of /aː/ responses. The effect of Precursor F2 ($\beta = -0.631$, $z = -6.000$, $p < .001$) indicates that fewer /aː/ responses were given for the precursor with the high than the low F2. As in Experiment 1, an effect of Load Condition was observed ($\beta = 0.365$, $z = 2.565$, $p = .01$), demonstrating more /aː/ responses were given in the high than the low load condition. There was no interaction between Load Condition and Precursor F2 ($p > .4$), indicating that the effect of precursor was not modulated by Load Condition. We observed an interaction between Vowel F2 and Precursor F2 ($\beta = 0.455$, $z = 2.728$, $p = .006$), suggesting that the higher the vowel F2, the smaller the effect of the precursor. This was most likely induced by a 'ceiling effect' (similar to Experiment 1) for vowels with a higher F2.

Cross-experiment analysis

Combined analyses across the two experiments were carried out (1) to test whether the cognitive load manipulation affected perception in both experiments to a comparable extent, and (2) to test individual variation between, on the one hand, the extent to which participants suffered from the load manipulation in their visual search accuracy, and, on the other hand, their load effect in target word categorization.

First, inspection of the effect sizes of the cognitive load effects in both experiments (with our coding, the models' estimates can be used as measures of effect size) seems to suggest a somewhat smaller load effect in Experiment 2 ($\beta = 0.365$) compared to Experiment 1 ($\beta = 0.437$). Note, however, that the target stimuli differed between the two experiments: a duration continuum in Experiment 1 vs. a spectral continuum in Experiment 2. Therefore, we restricted the analysis of the combined datasets to the shared vowel token from the continuum midpoints ("token 2").

A GLMM was fit to test this subset for effects of Load Condition (with low load coded as −0.5 and high load coded as +0.5), Experiment (with Experiment 1 coded as −0.5 and Experiment 2 coded as +0.5), and their interaction, with participant as random effect with by-participant random slopes for both fixed effects and their interaction (Barr et al., 2013). This model revealed an effect
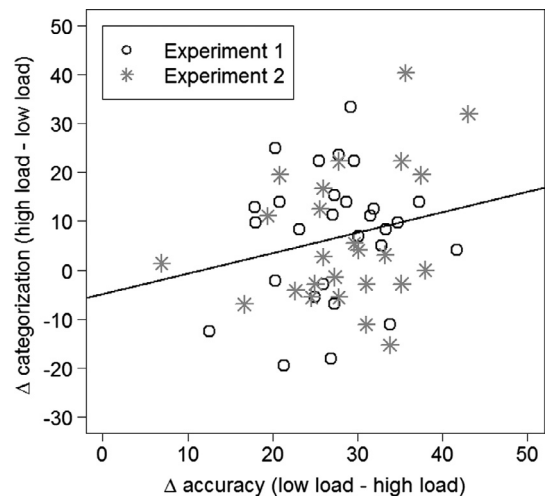

Fig. 5. Difference between load conditions in visual search accuracy scores ("Δ accuracy" in %; low load *minus* high load) plotted against the difference between load conditions in /aː/ categorization ("Δ categorization" in %; high load *minus* low load), for each participant in Experiment 1 (open circles) and Experiment 2 (star symbols). Note: only the data from trials with vowel token 2 were used, since only this token was shared between Experiment 1 and 2. The solid line gives the regression line.

of Load Condition ($\beta = 1.013$, $z = 4.067$, $p < .001$), but no significant effect of Experiment. Moreover, no interaction between Load Condition and Experiment was found. Thus, no evidence was found that would suggest a differential effect of cognitive load across the two experiments.

Second, for each participant, we calculated the difference in visual search accuracy between low and high load trials ("Δ accuracy"; low load accuracy *minus* high load accuracy; higher values indicate greater susceptibility to the load manipulation). We also calculated, for each participant, the difference in categorization between high and low load trials ("Δ categorization"; high load %/aː/ *minus* low load %/aː/; higher values indicate a greater effect of cognitive load on perception). Data from both experiments were combined to increase the sample size for the linear regression analysis ($N = 54$), but, again, only for the subset of shared target items (the "token 2" trials, i.e., the middle steps of the vowel continua). A simple linear regression was calculated to predict "Δ categorization" based on "Δ accuracy". As presented in Fig. 5, a significant linear regression equation was found ($b = 0.418$, $F(1, 52) = 4.030$, $p = .049$, *adjusted* $R^2 = 0.054$). This regression analysis suggests that as participants suffered more from the load manipulation in the visual search task, they also categorized the target words as more /aː/-like. However, a linear regression with only the accuracy scores from the high load condition as predictor did not reach significance, causing us to be cautious to make strong inferences based on this specific finding.

Discussion

The results from Experiment 2 mirror the results from Experiment 1. No evidence was found for an interaction between the contextual effect of the precursor's spectral

properties and the effect of cognitive load. However, once more a greater proportion of /a:/ responses was observed in the high load than in the low load condition. These results provide support for a 'shrinking of time' mechanism underlying effects of cognitive load on speech encoding. That is, independent of the spectral (or temporal; see Experiment 1) properties of the precursor, speech presented in the high load condition biased perception of subsequent target vowels towards /a:/. This could be interpreted such that under high cognitive load the precursors were perceived to have a perceptually faster speech rate.

In line with this argument, an analysis of the combined datasets from Experiment 1 and 2 did not reveal any evidence for differential effect sizes of the cognitive load effects in the two experiments. In addition, there was some evidence (*adjusted R^2* = 0.054) that those participants that performed worse in the visual search task under cognitive load also showed a greater proportion of /a:/ responses in target categorization. This suggests that individual variation in performance on the visual search task predicted participants' categorization patterns.

General discussion

The present study investigated two mechanisms that have been suggested to explain how cognitive load disrupts the encoding of fine phonetic detail. Specifically, it targeted potential effects of cognitive load on two acoustic context effects, namely rate normalization and spectral normalization. A 'noisy encoding' mechanism (Gordon et al., 1993; Mattys & Wiget, 2011) suggests that cognitive load disrupts speech encoding by modulating the signal-to-noise ratio in the perceptual system. That is, an increase in attentional demands is proposed to result in a reduction in the robustness of representation of the speech signal and/or an inadequate filtering of system noise, affecting the encoding of all phonetic cues in the signal. As a consequence, it predicted for our experiments that under increased cognitive load the temporal and spectral speech properties of a sentence context would be encoded in an impoverished form. This, in turn, would predict that their contextual influence on subsequent target sound perception would be attenuated. However, neither of the two experiments demonstrated a modulating influence of cognitive load on temporal or spectral context effects.

Although cognitive load did not modulate acoustic context effects (neither spectral nor temporal), we did find a consistent main effect of cognitive load on perception. That is, both experiments revealed that an increase in cognitive load during context presentation biased target perception towards /a:/. This observation is important for two reasons. First, it demonstrates that indeed our cognitive load manipulation was successful in causing a perceptually relevant effect. This shows that the lack of an influence of cognitive load on context effects was not simply caused by a weak cognitive load manipulation.

Second, and more interestingly, this finding can be interpreted as empirical support for a 'shrinking of time' mechanism, explaining cognitive load effects in speech perception. It has been argued (Block et al., 2010) and shown empirically (Casini et al., 2009) that increased attentional demands accelerate the perceived passing of time, hence decreasing the perceived duration of perceptual events. For the present experiments this would mean that high cognitive load increased the perceptual speech rate of the precursors which may have induced an additional temporal context effect of its own, above and beyond any contextual influences from the acoustic properties of the precursor. In line with this reasoning, we observed similar effect sizes of cognitive load in both experiments, irrespective of whether the actual manipulation of the precursor affected temporal or spectral characteristics of the speech signal. Since our target contrast could always be interpreted with regard to its perceived duration, the finding that cognitive load affected target perception equally in both experiments strongly supports the explanation that the rate of the precursors was perceptually speeded by the load manipulation. In addition we found some evidence that individual variation in susceptibility to our cognitive load manipulation predicted the individual effects of cognitive load in speech perception.

Furthermore, the present study extends our understanding of the 'shrinking of time' mechanism by using an implicit task. That is, where previous studies assessed time perception by means of explicit ratings (i.e., explicitly instructing participants to judge the duration of a specific time interval; Coull et al., 2004; Macar et al., 1994), our experiments may be said to assess implicit time perception, since participants were not explicitly pointed towards the duration of the precursor sentences. The only study that also used an implicit task, Casini et al. (2009), examined spoken stimuli with relatively short durations (e.g., single speech segments). Our findings demonstrate that longer stretches of speech (i.e., our precursors) are similarly susceptible to shrinking of time induced by increases in cognitive load. This shows that not only the perception of spoken durations (i.e., length of a single time interval) but also the perception of speech rate (i.e., ratio of syllables per time unit) is affected by cognitive load.

However, it should be noted that our findings in support of a 'shrinking of time' mechanism do not exclude more general effects of a 'noisy encoding' mechanism in other tasks or listening situations. The fact that cognitive load disrupts sub-lexical encoding leading to less robust low-level representations beyond the time domain is well attested in the literature (Mattys & Palmer, 2015; Mattys & Wiget, 2011; Mattys et al., 2013). Nevertheless, all these studies tested cognitive load effects on the perception of perceptually ambiguous speech (consonant continua; Mattys & Wiget, 2011) or speech in noise (Mattys & Palmer, 2015; Mattys et al., 2013). In contrast, in the current study, the speech presented during the dual task (i.e., the precursor) was (although manipulated) unambiguous and noise-free. It may be speculated that cognitive load effects on the encoding of fine phonetic detail are rather subtle, and are only visible when the perceptual system is already challenged by low quality phonetic input. Thus, the present findings serve to extend our understanding of the situations in which speech representations are susceptible to modulating influences of cognitive load.

Moreover, the outcomes of the current study speak to another debate in speech perception on the temporal ordering of different perceptual processes. Specifically, the present results corroborate the view that acoustic context effects operate at an early locus in perception (Bosker, 2016b; Reinisch & Sjerps, 2013; Sjerps & Reinisch, 2015; Toscano & McMurray, 2015). Our results argue for a temporal ordering of effects such that acoustic context effects precede any modulating effects of cognitive load. This proposal is in line with a recent study by Mitterer and Mattys (2016). They argue that cognitive load does not necessarily or solely reduce the fidelity of early perceptual processes. Instead, they suggest that cognitive load affects speech processing by competing for resources in working memory. That is, the locus of effects of cognitive load is working memory. Considering that acoustic context effects are early and of a general acoustic nature, they may be assumed to be independent of working memory resources, explaining the absence of modulating influences of cognitive load.

In addition, however, given that cognitive load itself appears to trigger another type of context effect by speeding up perceived passing of time, our data support a two-stage model of context effects in speech perception. At the first stage, an automatic general auditory mechanism is operating unaffected by attentional modulation. Examples of such a mechanism are auditory contrast (spectral contrast; Laing et al., 2012; durational contrast; Wade & Holt, 2005) or, specifically concerning rate normalization, neural entrainment to the syllabic rate of speech (Bosker, 2016b; Peelle & Davis, 2012). This first stage operates early in perception (Reinisch & Sjerps, 2013; Sjerps & Reinisch, 2015; Toscano & McMurray, 2015), can be triggered by non-speech contexts (Bosker, 2016b; Laing et al., 2012; Wade & Holt, 2005), and is relatively robust against talker changes (Bosker, 2016a; Newman & Sawusch, 2009). The present study adds to this literature by revealing that it is robust against changes in attentional demands (see also Sjerps, McQueen, & Mitterer, 2012).

At a later stage/point in time, however, higher-level influences come into play when listeners make a decision on their classification of the target sounds. This would involve a comparison between a target sound and its expected realization given a certain context (see, e.g., Assgari & Stilp, 2015; Barreda, 2012; Glidden & Assmann, 2004; Johnson, 1990; Johnson, Strand, & D'Imperio, 1999; Nusbaum & Magnuson, 1997; Nusbaum & Morin, 1992). Support for this interpretation comes from the finding that foreign-accented speech as well as speech that contains fast-speech processes such as segmental reductions and deletions is implicitly perceived as faster (Bosker & Reinisch, 2015; Reinisch, 2016a), possibly due to greater listening effort. In the current experiments, the online processing of the precursors was unaffected by increased cognitive load (at least at, or before, the level where acoustic context effects influenced perception). We can thus speculate that the 'shrinking of time' may only have affected listeners' memory of the precursor sentences. Thus, the shift in the categorization boundary of the target continuum may have been a result of cognitive rather than perceptual adjustments (e.g., Barreda, 2012; Nusbaum & Morin, 1992).

Further research is needed, however, to shed more light on this issue

In conclusion, the present study suggests that cognitive load induces a perceptual increase in perceived speech rate but does not modulate acoustic context effects in speech perception, supporting a mechanism that perceptually shrinks time. Given the lack of evidence for modulation of acoustic context effects, we propose that acoustic context effects occur independently and likely earlier during processing than modulating effects of cognitive load. This could be explained by a two-stage model of context effects. That is, at an initial stage in perception, context-dependent processing takes place independent of any concurrent attentional demands on the cognitive system. Later, higher-level influences such as the perceived acceleration of time, induced by cognitive load, may still cause additional influences on listeners' decisions.

Acknowledgments

References

Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. *The Journal of the Acoustical Society of America, 138*(5), 3023–3032.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Barreda, S. (2012). Vowel normalization and the perception of speaker changes: An exploration of the contextual tuning hypothesis. *The Journal of the Acoustical Society of America, 132*(5), 3453–3464.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*(3), 330–343.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [computer program]: Version 5.4.12, retrieved from <http://www.praat.org/>.

Bosker, H. R. (2016a). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* http://dx.doi.org/10.1037/xlm0000381.

Bosker, H. R. (2016b). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception & Psychophysics.* http://dx.doi.org/10.3758/s13414-016-1206-4 (OnlineFirst).

Bosker, H. R., & Reinisch, E. (2015). Normalization for speechrate in native and nonnative speech. In M. Wolters, J. Livingstone, B. B., R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th international congress of phonetic sciences 2015 [ICPhS XVIII]*, Glasgow.

Casini, L., Burle, B., & Nguyen, N. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition, 112*(2), 318–322.

Coull, J. T., Vidal, F., Nazarian, B., & Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science, 303*(5663), 1506–1508.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America, 102*(3), 1891–1897.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *The Journal of the Acoustical Society of America, 85*(5), 2154–2164.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452–465.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). Cognitive noise is also noise: The impact of attention on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics, 72*, 1522–1532.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110.

Gerrits, E. (2001). *The categorisation of speech sounds by adults and children*. Utrecht University.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology, 25*, 1–42.

Glidden, C. M., & Assmann, P. F. (2004). Effects of visual gender and frequency shifts on vowel category judgments. *Acoustics Research Letters Online, 5*(4), 132–138.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., & Simon, J. Z. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron, 77*(5), 980–991.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*(2), 137–146.

Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America, 125* (6), 3983–3994.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *The Journal of the Acoustical Society of America, 88*(2), 642–654.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359–384.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *The Journal of Neuroscience, 30*(2), 620–628.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America, 29*(1), 98–104.

Laing, E. J., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology, 3*, 1–9.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix* japonica). *The Journal of the Acoustical Society of America, 102*(2), 1134–1140.

Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition, 22*(6), 673–686.

Mattys, S. L., Barden, K., & Samuel, A. G. (2013). Impaired speech recognition under a cognitive load: Where is the locus? *Proceedings of Meetings on Acoustics* (Vol. 19, pp. 1–6). Acoustical Society of America.

Mattys, S. L., & Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America, 137*(3), 1464–1472.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65*(2), 145–160. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0749596X1100043X>.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485* (7397), 233–236.

Mitterer, H., & Mattys, S. L. (2016). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception & Psychophysics*. http://dx.doi.org/10.3758/s13414-016-1195-3 (OnlineFirst).

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics, 37*(1), 46–65.

Nooteboom, S., & Cohen, A. (1984). *Het proces van spreken en verstaan, een nieuwe inleiding in de experimentele fonetiek*. Assen, The Netherlands: Van Gorcum.

Nusbaum, H. C., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109–132). Academic Press.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 113–134). OHM Publishing.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech, 3*(1), 11–17.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413–425.

R Development Core Team (2012). R: A Language and Environment for Statistical Computing.

Reinisch, E. (2016a). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics, 78*(4), 1203–1217. http://dx.doi.org/10.3758/s13414-016-1067-x.

Reinisch, E. (2016b). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics, 37*, 1397–1415. http://dx.doi.org/10.1017/S0142716415000612.

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics, 41*(2), 101–116.

Sinnott, J. M., Brown, C. H., & Borneman, M. A. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. *Perception & Psychophysics, 60* (6), 1032–1043.

Sjerps, M. J., McQueen, J. M., & Mitterer, H. (2012). Extrinsic normalization for vocal tracts depends on the signal, not on attention. In *Proceedings of INTERSPEECH 2012* (pp. 394–397).

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics, 73*(4), 1195–1215.

Sjerps, M. J., & Reinisch, E. (2015). Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 710–722.

Stilp, C. E., Alexander, J. M., Kiefte, M., & Kluender, K. R. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception, & Psychophysics, 72*(2), 470–480.

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience, 30*(5), 529–543.

van Heuven, V., Van Houten, J., & De Vries, J. (1986). De perceptie van Nederlandse klinkers door Turken. *Spektator, 15*, 225–238.

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 67*(6), 939–950.

Welch, T. E., Sawusch, J. R., & Dent, M. L. (2009). Effects of syllable-final segment duration on the identification of synthetic speech continua by birds and humans. *The Journal of the Acoustical Society of America, 126*(5), 2779–2787.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language, 36*, 165–187.