Journal of Memory and Language 94 (2017) 166–176

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

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

Article history:
Received 2 March 2016
revision received 21 November 2016
Available online 24 December 2016

Keywords:
Cognitive load
Acoustic context
Rate normalization
Spectral normalization

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.

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).

http://dx.doi.org/10.1016/j.jml.2016.12.002

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