Hearing Research 323 (2015) 81–90

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

The pupil response reveals increased listening effort when it is difficult to focus attention

Thomas Koelewijn [a,*], Hilde de Kluiver [a], Barbara G. Shinn-Cunningham [b], Adriana A. Zekveld [a,c], Sophia E. Kramer [a]

[a] Section Ear & Hearing, Department of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

[b] Department of Biomedical Engineering, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, USA

[c] Linnaeus Centre HEAD, Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

ARTICLE INFO

Article history:
Received 22 October 2014
Received in revised form
5 February 2015
Accepted 16 February 2015
Available online 27 February 2015

ABSTRACT

Recent studies have shown that prior knowledge about where, when, and who is going to talk improves speech intelligibility. How related attentional processes affect cognitive processing load has not been investigated yet. In the current study, three experiments investigated how the pupil dilation response is affected by prior knowledge of target speech location, target speech onset, and who is going to talk. A total of 56 young adults with normal hearing participated. They had to reproduce a target sentence presented to one ear while ignoring a distracting sentence simultaneously presented to the other ear. The two sentences were independently masked by fluctuating noise. Target location (left or right ear), speech onset, and talker variability were manipulated in separate experiments by keeping these features either fixed during an entire block or randomized over trials. Pupil responses were recorded during listening and performance was scored after recall. The results showed an improvement in performance when the location of the target speech was fixed instead of randomized. Additionally, location uncertainty increased the pupil dilation response, which suggests that prior knowledge of location reduces cognitive load. Interestingly, the observed pupil responses for each condition were consistent with subjective reports of listening effort. We conclude that communicating in a dynamic environment like a cocktail party (where participants in competing conversations move unpredictably) requires substantial listening effort because of the demands placed on attentional processes.

1. Introduction

Having a conversation with a good friend at a party can be relatively easy if you know where and when he or she is going to talk to you (e.g., Kitterick et al., 2010). On the other hand, talking at the same party with a group of people whom you do not know well and who are dancing or moving around feels much more effortful. Although multiple studies show that prior knowledge about where, when, and who is talking has a positive effect on speech recall performance (e.g., Ihlefeld and Shinn-Cunningham, 2008; Kitterick et al., 2010), there is little evidence that this information affects cognitive load during speech processing.

We showed in a previous study (Koelewijn et al., 2014) that dividing attention over two streams of information instead of focusing on one increases cognitive load. According to the 'load theory of selective attention' (Lavie et al., 2004), high cognitive load decreases performance, an effect observed in our study and in prior research (Best et al., 2010). We concluded that the amount of allocated attentional recourses affects cognitive load. If these attentional resources are deployed effectively, this should lead to better segregation of target information from background information and thus better performance (Broadbent, 1958). Effective early filtering should ease later semantic processing by reducing the amount of conflicting information vying for resources (Rönnberg et al., 2013), thereby reducing the total cognitive load. This was not addressed in our previous study (Koelewijn et al., 2014), where we only investigated the amount of cognitive resources needed to process two streams of information compared to one and not how effectively attentional processes could use

* Corresponding author. VUmc – Ear & Hearing, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands. Tel.: +31 20 44 40900.
E-mail address: t.koelewijn@vumc.nl (T. Koelewijn).

http://dx.doi.org/10.1016/j.heares.2015.02.004

available cues to facilitate target-masker segregation processes. For effective early filtering, listeners must be able to access relevant, salient cues that distinguish target from masker to enable attention to be properly focused on the target. In the current study, we investigate how the features location, speech onset, and voice (and other speech characteristics) of a talker affect speech intelligibility and listening effort.

Kidd et al. (2005) showed that in a complex listening task when there were two distractor talkers, prior knowledge about *where* the target speech is presented has a positive effect on performance. This effect was replicated by Kitterick et al. (2010), who simulated a complex listening environment in order to create challenges like those that arise at a cocktail party. The effects of uncertainty of speech location, speech onset, and target talker on speech perception were investigated by determining the benefits of constraining these three parameters during speech reception threshold (SRT) tasks. Target phrases were masked by at least 12 distracting phrases within each trial. Constraining where the target talker was located yielded a modest benefit of 1.0 dB in SRT when the target phrases and the masking phrases had different onset times relative to one another. When one of the masking phrases had an onset time similar to that of the target phrase, the benefit of location information reached 5.1 dB. In other words, the location information became more relevant when a distracting sound was presented at the same time. In a study by Best et al. (2007), visually guided attention was directed towards the location of a talker or a particular birdsong. Their results showed that knowledge about where a target is located improves its identification when presented with similar distractors.

The effect on speech perception of knowing *when* someone is going to speak has not been studied much. Best et al. (2007) showed that visually cueing the target onset had little effect on the ability to attend to and recall a spoken digit stream and a modest effect for birdsongs. Gatehouse and Akeroyd (2008) showed a small performance benefit for hearing-impaired listeners when the onset of a word was preceded by a visual cue. Kitterick et al. (2010) also showed a small effect of making speech target onset times more predictable. Thus, providing temporal information yields small benefits for the behavioral ability to attend to and recall speech.

Finally, Kitterick et al. (2010) showed that constraining *who* is going to talk affects speech intelligibility. In their study, the target talker was either fixed or randomly selected from of a group of talkers. When the same talker uttered several target phrases, participants were able to perform the task under less favorable listening conditions (lower signal-to-noise ratios, SNRs) than when the target phrases were uttered randomly by one of the talkers. The results suggest that prior knowledge about who is going to talk benefits speech processing. This is in line with the idea that familiar voices are more intelligible than novel voices (e.g., Nygaard and Pisoni, 1998) and that content from learned voices is better encoded in or recalled from memory (e.g., Martin et al., 1989). Other studies (Brungart and Simpson, 2004; Brungart et al., 2001) have also shown that prior knowledge of the vocal characteristics of either the target talker or a distracting talker improves performance in speech intelligibility tasks. In all, prior knowledge that allows focusing of attention on who is going to speak, and where and when this is going to occur, enhances speech intelligibility.

There is a relationship between cognitive processes such as attention, and the pupillary response (Beatty, 1982). Increased cognitive task demands reliably induce a larger pupil dilation response, allowing task-evoked pupillary responses to be used as a reliable and valid measure of cognitive processing load (Just et al., 2003; Kahneman and Beatty, 1966). Consequently, the task-evoked pupillary response quantifies listening effort in auditory tasks (Hyönä et al., 1995; Kramer et al., 1997). Generally, when a task requires more processing load in the same time interval, mean pupil dilation is larger when the task is being performed (Granholm and Verney, 2004). Additionally, in this same time window one can measure both the peak pupil dilation, which is thought to represent the maximum processing load, and peak latency, which is associated with processing time (Zekveld et al., 2011). The mean and peak pupil dilation are measured relative to a baseline, typically defined by the mean pupil diameter during a period of time in which no task-related processing occurs (e.g., over a time window one second prior to the onset of the target stimulus). In the current study, we analyzed all of these pupil measures, as they provide insight into how attention affects overall cognitive load (mean pupil dilation), maximum cognitive load (peak pupil dilation), processing speed of higher cognitive processes (peak latency), and overall task engagement (baseline), as explained below.

Pupil diameter is tightly linked to the activity of the Locus Coeruleus (LC) (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010). The noradrenergic system of the LC (LC-NE) is associated with various psychological processes, including attention. The activity of the LC-NE seems to exhibit two modes of function: phasic and tonic. During task performance, the phasic mode is associated with large responses to task-related events and low baseline firing rate of the LC-NE. The tonic mode is associated with high baseline activity of the LC-NE and a lack of phasic responses. The adaptive gain theory of Aston-Jones and Cohen (2005) proposes that the phasic mode is driven by optimization of performance (exploitation) and task engagement, whereas the tonic mode favors exploration of the environment, greater distractibility (sensitivity to task-irrelevant stimuli), and task disengagement. Rajkowski et al. (1994) investigated the relationship between the baseline pupil diameter and the LC-NE mode. The phasic and tonic modes were marked by relatively small and large baseline diameter values, respectively. It has been suggested that the task-evoked pupillary response corresponds to the phasic activity of the LC-NE, whereas the baseline pupil diameter corresponds to the tonic activity (Gilzenrat et al., 2010; Meer et al., 2010).

The main aim of the current study was to examine whether or not location, speech onset, and target talker uncertainty have an effect on the pupil response during speech perception tasks. Target location, onset, and talker variability were manipulated in three separate experiments. During these experiments, participants with normal hearing were presented with auditory sentences in fluctuating noise. Participants were asked to focus attention and repeat back target sentences while simultaneously ignoring distracting stimuli. We tested the hypothesis that allowing attention to focus on location, onset, or talker voice, would make it easier for listeners to filter out irrelevant information during early processing. Consequently, processing load would be reduced, as reflected by a smaller pupil dilation response and increased performance. In addition, participants gave subjective effort ratings after each task to allow us to evaluate how well cognitive load (specifically that related to attentional processes) reflects subjective listening effort.

2. Experiment 1: location uncertainty

In Experiment 1, we investigated the effect of location uncertainty on speech intelligibility and the pupil response (dilation, latency, baseline). We used a design similar to that employed in a previous study that examined the effect of divided attention on cognitive load (Koelewijn et al., 2014). In the current experiment, the location of the target speech was either fixed (location-fixed) during a block, by presenting sentences to the same ear, or varied (location-random) across trials, by randomly presenting the sentence to the left or right ear. We hypothesized that in the location-