The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load

Adriana A. Zekveld [a,b,*], Dirk J. Heslenfeld [c], Ingrid S. Johnsrude [b,d,e], Niek J. Versfeld [a], Sophia E. Kramer [a]

[a] Section Audiology, Dept. of Otolaryngology-Head and Neck Surgery and EMGO Institute for Health and Care Research, VU University Medical Center, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands
[b] Department of Behavioural Sciences and Learning, Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Linköping University, SE-581 83 Linköping, Sweden
[c] Department of Psychology, VU University, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands
[d] Department of Psychology, Queen's University, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada
[e] The School of Communication Sciences and Disorders and The Brain and Mind Institute, Natural Sciences Centre, Western University, London, Ontario N6A 5B7, Canada

ARTICLE INFO

ABSTRACT

An important aspect of hearing is the degree to which listeners have to deploy effort to understand speech. One promising measure of listening effort is task-evoked pupil dilation. Here, we use functional magnetic resonance imaging (fMRI) to identify the neural correlates of pupil dilation during comprehension of degraded spoken sentences in 17 normal-hearing listeners. Subjects listened to sentences degraded in three different ways: the target female speech was masked by fluctuating noise, by speech from a single male speaker, or the target speech was noise-vocoded. The degree of degradation was individually adapted such that 50% or 84% of the sentences were intelligible. Control conditions included clear speech in quiet, and silent trials.

The peak pupil dilation was larger for the 50% compared to the 84% intelligibility condition, and largest for speech masked by the single-talker masker, followed by speech masked by fluctuating noise, and smallest for noise-vocoded speech. Activation in the bilateral superior temporal gyrus (STG) showed the same pattern, with most extensive activation for speech masked by the single-talker masker. Larger peak pupil dilation was associated with more activation in the bilateral STG, bilateral ventral and dorsal anterior cingulate cortex and several frontal brain areas. A subset of the temporal region sensitive to pupil dilation was also sensitive to speech intelligibility and degradation type. These results show that pupil dilation during speech perception in challenging conditions reflects both auditory and cognitive processes that are recruited to cope with degraded speech and the need to segregate target speech from interfering sounds.

© 2014 Elsevier Inc. All rights reserved.

Introduction

In the variable and noisy listening conditions of everyday life, accurate speech perception can be cognitively demanding, particularly for those with hearing loss. In recent years, researchers have increasingly focused on validating methods to quantify the processing load required during speech perception. These methods include subjective rating scales and the pupil response (Kramer et al., 2013; Mackersie and Cones, 2011; Piquado et al., 2010; Tun et al., 2009; Zekveld et al., 2010). However, it is unknown to which extent these behavioral and physiological measures rely on a common neural substrate. In the present study, we investigated the neural correlates of pupil dilation evoked by listening to degraded speech.

The task-evoked pupil response has been demonstrated in numerous studies (Beatty, 1982; Engelhardt et al., 2010; Granholm et al., 1996; van der Meer et al., 2010). It appears to reflect activation of the locus coeruleus (LC) norepinephrine (LC-NE) system, which plays an important role in controlling autonomic functions (Joshi et al., 2013; Samuels and Szabadi, 2008a,b). The LC has dense projections to many cerebral regions via the excitatory neurons of the thalamus. It additionally has inhibitory effects on the parasympathetic nervous system (Samuels and Szabadi, 2008a). The LC-NE system has been associated with several cognitive functions such as memory, attention, reward anticipation and decision-making (Einhäuser et al., 2010; Laeng et al., 2012).

Brain regions that are associated with the pupil response and the LC are the orbitofrontal cortex, ventral and dorsal anterior cingulate (AC) cortex, medial prefrontal cortex, dorsal pons and bilateral frontal insula (Critchley et al., 2005). These areas have been associated with evaluation of rewards (orbitofrontal cortex) and costs (e.g. errors in performance, task difficulty and conflicts in processing) (Critchley et al., 2005). Furthermore, the LC-NE system has been associated with

* Corresponding author at: Otolaryngology-Head and Neck Surgery, Section Audiology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. Fax: +31 20 444 2033.
E-mail address: aa.zekveld@vumc.nl (A.A. Zekveld).

attentional processes, because LC activity depends on behavioral state (e.g., sleep–wake cycle) and is modulated by stimuli that result in orienting responses (Aston-Jones et al., 1999; Corbetta et al., 2008).

Johnstone et al. (2007) applied a task involving processing of emotional pictures and observed that individuals who had larger pupil dilation while reappraising negative stimuli had a relative decrease in activation in the bilateral insula, left amygdala, and in a large midbrain/thalamic cluster during reappraisal. These areas are associated with the generation and experience of negative emotion (Johnstone et al., 2007). Sterpenich et al. (2006) observed that LC activation during the retrieval of emotional memories was related to the peak pupil dilation during encoding. Additionally, the LC was functionally connected more tightly with the amygdala when the correctly remembered images were encoded in an emotional context than when they were encoded in a neutral context (Sterpenich et al., 2006).

Recently, two studies demonstrated a relationship between activation in the LC and several additional subcortical and cortical regions and the pupil response to cognitive load. In two test sessions, Alnaes et al. (2014) measured the pupil response or brain activation using functional magnetic resonance imaging (fMRI) while subjects performed a visual object tracking task which has been associated with attentional load. Activation in the LC was associated with pupil size, as well as activation in the inferior precentral gyrus, precentral sulcus, frontal eyefield, superior frontal gyrus, and several parietal, occipital and subcortical areas. The authors interpret the results as indicating that the pupil response provides a reliable index of attentional effort. Murphy et al. (2014) concurrently measured fMRI activation and the pupil response to an oddball task. LC activation was associated with the pupil response, and additionally with stimulus identity (target versus standard stimuli). The authors conclude that the pupil diameter could be a proxy measure for LC activity.

These studies indicate that pupil dilation is associated with activation in several cortical and subcortical regions during a variety of emotional and cognitive tasks not involving perception of speech. It is unknown whether pupil dilation evoked by speech comprehension in difficult listening conditions activates the same areas. The areas that have been suggested to be involved in effortful speech perception include the left inferior frontal gyrus and the AC cortex (Davis and Johnsrude, 2003, 2007; Erb and Obleser, 2013; Obleser and Kotz, 2010; Scott et al., 2004; Wild et al., 2012; Zekveld et al., 2006, 2012). Sensitivity of the pupil response to effortful processing during speech perception would be reflected in an association between activation in these areas and the task-evoked pupil response during listening. To examine this hypothesis, we concurrently assessed brain activity using fMRI and the peak pupil dilation while subjects perceived speech in several conditions that induced different amounts of processing load. We presented degraded speech at two intelligibility levels (50% and 84% correct sentence perception) using three types of degradation that were created by imposing either fluctuating noise or a single-talker masker or by noise-vocoding the target speech. Speech masked by a single-talker evokes a larger peak pupil dilation than speech masked by fluctuating noise (Koelewijn et al., 2012a,b) and pupil dilation is larger when the stimulus is more degraded (Zekveld and Kramer, 2014; Zekveld et al., 2010). Previous neuroimaging studies showed that listening to speech masked by speech results in more activation in the superior temporal gyrus (STG), superior temporal sulcus and the supratemporal planes as compared to speech masked by noise (Scott and McGettigan, 2013; Scott et al., 2004, 2009), which may suggest that the to-be-ignored signal is processed to some extent (Scott and McGettigan, 2013).

We performed a conjunction analysis to assess the overlap in brain regions associated with the peak pupil dilation and those associated with the various speech-listening conditions. Activation overlap between these regions may indicate that the neural processes associated with effortful listening are also reflected by the pupil dilation. This would support the interpretation of the pupil response as reflecting processing effort imposed by decreased quality of speech, or by increased segregation demands due to the different types of degradation (speech masked by interfering speech, by fluctuating noise, or noise-vocoded). We expected that such regions would include those that have been associated with effortful speech perception, like the left inferior frontal gyrus (Davis and Johnsrude, 2003, 2007; Erb and Obleser, 2013; Obleser and Kotz, 2010; Scott et al., 2004; Wild et al., 2012; Zekveld et al., 2006, 2012).

Materials and methods

Participants

Eighteen normal-hearing young adults (9 women and 9 men) participated. Age ranged from 19 to 33 yrs, with a mean of 24.2 yrs ($SD = 3.9$ yrs). The pupil data acquired in the fMRI session of one participant were of insufficient quality and therefore this subject's data were omitted from analysis, resulting in a sample size of 17. Participants were recruited from among students and employees of VU University and VU University Medical Center, and via advertisements. All participants were native Dutch speakers, and right-handed on the Dutch "Classificatie van links- en rechtshandige proefpersonen" ("Classification of left- and right-handed subjects") (Van Strien, 1992). Participants all had normal or corrected-to-normal vision when wearing glasses or contact lenses and were able to read text with font size of 6 points or smaller, as determined by screening with a near-vision test equivalent to a visual acuity Snellen chart (Bailey and Lovie, 1980). Color vision was screened with Ishihara plates (Ishihara, 1989); this was for another task completed by these participants that is not reported here. The exclusion criteria were the following: pure-tone air-conduction hearing threshold(s) in one or both ears exceeding 20 dB HL at the octave frequencies between 500 and 8000 Hz, any middle-ear abnormalities as assessed with tympanometry, dyslexia or other reading problems, claustrophobia, epilepsy, having a history of a neurological or psychiatric disease, having metal in the body that would preclude an MRI scan, and being pregnant or on medication. All participants had normal hearing; the mean pure-tone hearing threshold at 500, 1000, 2000, and 4000 kHz, averaged over both ears, was 4.0 dB HL (SE = 2.4 dB) and all had normal tympanograms. All participants provided written informed consent in accordance with the Ethics Committee of the VU University Medical Center.

Tests and procedures

Participants were tested in two sessions that took place on different days. On average, session 2 was 36 days after session 1 (ranging between 1 and 66 days). In the first session, behavioral and pupillometric data were acquired. This session included tests to estimate an appropriate signal-to-noise ratio (SNR) for the imaging study. In the second session, fMRI and pupillometric data were acquired concurrently. We will first summarize the procedures, and then describe these in detail.

General procedures

Session 1 took place in a sound-attenuating room. The test session started with pure-tone audiometry, near vision and color vision screening. Then participants filled in the handedness questionnaire. We additionally measured six conditions of the speech reception threshold (SRT) test (Plomp and Mimpen, 1979). Conditions conformed to a two (intelligibility levels: 50% and 84% correct sentence perception) by three (degradation types: fluctuating noise masker, single-talker masker, noise-vocoded speech) within-subjects design, with the addition of a control condition that consisted of clear, undegraded, sentences (speech in quiet). In the fluctuating noise and single-talker masker conditions, the speech-to-noise ratio was adaptively varied to estimate the SNR