Pupil size tracks semantic ambiguity as well as noise

Mason Kadem, The University of Western Ontario

Supervisor: Ingrid Johnsrude, The University of Western Ontario

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Neuroscience

© Mason Kadem 2019

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Cognitive Neuroscience Commons

Recommended Citation

https://ir.lib.uwo.ca/etd/6426
Abstract

Effortful listening is experienced by listeners when speech is hard to understand because it is degraded or masked by environmental noise. Pupillometry (i.e., measure of pupil size) can detect effortful listening: pupil size increases when speech is degraded compared to when it is clear. However, the pupil responds to a range of cognitive demands, including linguistic challenges such as syntactic complexity. Here I investigate whether it responds to the need to disambiguate words with more than one meaning, such as ‘bark’ or ‘bank’. Semantic ambiguity is common in English, and previous work indicates that it imposes a processing load. We combine this with an acoustic challenge in a factorial design so the pupil response to these two types of challenge can be directly compared. I found main effects of noise and semantic ambiguity on the pupillary area, indicating that pupil dilation can reflect processes associated with semantic disambiguation as well as noise. Pupil size reflect demands imposed by ambiguity both in the acoustic form of words (i.e. due to degradation) and in word meaning.

Keywords

Pupil; Cognitive effort; Processing demands; Semantic ambiguity; Background noise
Layman Abstract

It is difficult to listen to speech when it is either spoken unclearly, or if there is background noise affecting our ability to hear it. This difficulty in hearing requires higher mental activity called cognitive load, or more work for our brain when it comes to understanding speech. The pupils of our eyes change its size depending on how hard our brain is working, and the size therefore tells us something about the level of cognitive load. The way we can measure pupil size is through a method called pupillometry. Pupillometry can measure a change in pupil size that relates to whether speech is heard clearly or is noisy and degraded. While this is a useful method, the pupil can change its size when listening to speech for reasons that are not related to the quality of the signal. Cognitive factors involved in comprehension also make our brains work hard, such as when we can hear a sentence containing words with more than one meaning, such as “The shell was fired towards the tank”. The correct meaning of the bolded words must be deduced based on the other words in the sentence. In this thesis, our goal was to compare pupillometry measures when a listener has to work to understand a sentence, when words are hard to hear, and when their meaning has to be deduced from context.
Co-Authorship Statement

MK helped design the study, collected, analyzed and disseminated the data, and wrote and edited the manuscript. BH assisted with dissemination, analysis, design and edited the manuscript. JR provided the British stimuli. ISJ conceived and designed the study, contributed to data dissemination, re-recorded the stimuli and edited the manuscript.
Acknowledgments

Thank you, Ingrid, for fueling my research dreams and supporting me
Thank you, Björn, for helping me develop my programming and technical tree
Jody Culham and Brian Corneil, for making the thesis as good as it can be
Merci to friends who helped me with this degree
Special thanks to these three: my mom, Sahar, and my brothers, Sarmad and Fawazy
for the important roles they have played in my life. Now for my wife —
Baraa, to you I dedicate this thesis, for $10^{18}$ reasons.
For the one’s that need us, the little wee ones, Jack and Emmi, Love you three tons.
Table of Contents

Abstract ................................................................................................................................. ii
Layman Abstract .................................................................................................................. iii
Co-Authorship Statement .................................................................................................... iv
Acknowledgments ............................................................................................................... v
Table of Contents ............................................................................................................... vi
List of Figures ..................................................................................................................... vii
List of Appendices .............................................................................................................. viii
Chapter 1 ............................................................................................................................ 1
  1.1 Listening Challenges ........................................................................................................ 1
  1.2 Processing Demands ....................................................................................................... 2
    1.2.1 Perceptual demands ............................................................................................... 2
    1.2.2 Linguistic demands ............................................................................................. 3
  1.3 Semantic Ambiguity ....................................................................................................... 4
  1.4 Neuroimaging correlates of semantic ambiguity .......................................................... 6
  1.5 Pupillometry ................................................................................................................ 9
    1.5.1 Physiology of the pupillary response .................................................................... 10
  1.6 Microsaccades ............................................................................................................ 11
  1.7 Rationale for the current study .................................................................................... 12
Chapter 2 ............................................................................................................................ 14
  2.1 Participants .................................................................................................................. 14
  2.2 Auditory stimuli and task ............................................................................................ 14
  2.3 Procedure and data recording ...................................................................................... 16
  2.4 Data analysis ............................................................................................................... 18
Chapter 3 ............................................................................................................................ 21
  3.1 Experiment 1 ............................................................................................................... 21
    3.1.1 Semantic relatedness task .................................................................................... 21
    3.1.2 Pupillometry ........................................................................................................ 21
  3.2 Experiment 2 ............................................................................................................... 23
    3.2.1 Semantic relatedness task .................................................................................... 23
    3.2.2 Pupillometry ........................................................................................................ 24
  3.3 Pooling Experiment 1 and 2 ....................................................................................... 25
  3.4 Microsaccade results ................................................................................................... 26
Chapter 4 ................................................................................................................................ 28
  4.1 Behavioural effects ...................................................................................................... 28
  4.2 Pupillary effects ......................................................................................................... 29
  4.3 Microsaccade effects .................................................................................................. 29
  4.4 Pupil dilation and listening effort ............................................................................... 31
  5 Conclusion ....................................................................................................................... 34
References ........................................................................................................................... 35
Appendices .......................................................................................................................... 41
Appendix A: Ethics approval ............................................................................................... 41
Appendix B: Letter of information and consent ................................................................. 42
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Neuroimaging correlates of semantic ambiguity and noise.</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Experimental design and results of semantic relatedness task.</td>
<td>16</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Semantic relatedness results (Experiment 1)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Pupil dilation results (Exp 1).</td>
<td>22</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Semantic relatedness results (Experiment 2)</td>
<td>24</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Pupil dilation results (Exp 2).</td>
<td>25</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Results for microsaccade analysis.</td>
<td>27</td>
</tr>
</tbody>
</table>
List of Appendices

Appendix A: Ethics approval

Appendix B: Letter of Information and consent
Chapter 1

1 Introduction

Cognitive processes compensate for decrements in audibility and intelligibility due to hearing loss (Whitson et al., 2018) and so listening to speech in noise can be cognitively effortful (Wendt et al. 2016, Johnsrude & Rodd, 2016). Pupillometry offers a methodological means for interrogating factors that affect speech reception, as 50 years of research show that the pupil is a reliable marker for effort. However, when using pupillometry measures as a metric of effort, it is important to consider that the pupil responds to a broad range of cognitively demanding processes. A common feature of speech is semantic ambiguity. We know semantic ambiguity increases processing load; yet, we do not know whether the pupil responds to semantic ambiguity. This thesis examines how semantic ambiguity and background noise affect the behavioural and physiological (pupillometry) indices of processing load during speech reception. A literature review follows of listening challenges and pupillometry.

1.1 Listening Challenges

Hearing loss is a common impairment and has consequences for cognitive processing (Ciorba et al., 2012). Cognitive processes compensate when audibility fails (Whitson et al., 2018), thus listening to speech in noise can feel effortful for affected individuals (Wendt et al. 2016, Johnsrude & Rodd, 2016). The increased recruitment of cognitive resources for understanding speech also limits those available for other tasks (Effortfulness hypothesis; Rabbitt, 1968; Wingfield et al., 2005), increases fatigue (Hornsby, 2013) and may impair quality of life (Alfakir et al., 2015).

After a hearing prosthesis (hearing aid or cochlear implant) is fitted, hearing abilities are assessed by evaluating the individual’s ability to identify words spoken in noise (e.g., degraded sentences) before and after prosthesis placement (Shanks
et al., 2007). Two individuals can reach the same level of word identification performance, yet one individual may achieve that level of intelligibility with minimal perceived effort while the other individual may experience a greater effort. The greater the perceived effort, the more dissatisfied an individual may be with their hearing device. Accounting for effort as well as intelligibility is necessary during clinical testing to accurately predict how well each individual will do after hearing-aid fitting.

The concept of ‘listening effort’ is gaining in popularity in clinical circles and may help explain hearing outcomes. “Listening effort” is a function of the resources required for processing speech, and relies on both the challenge introduced by the stimulus (e.g., background noise, sentence complexity) and the cognitive abilities possessed by the listener (e.g., cognitive processes recruited to cope with listening challenges) (Johnsrude and Rodd, 2015). The perception of effort reflects a variety of cognitive processes that are recruited to help understand speech.

There is growing interest in characterizing effortful listening. However, listening effort tends to be interpreted as reflecting cognitive demands associated with acoustic degradation as a consequence of peripheral hearing loss. However, speech utterances themselves may impose cognitive demands, as people work to understand what is said, and these combine with the acoustic demands to increase cognitive load.

1.2 Processing Demands
1.2.1 Perceptual demands
Perceptual demands result when the quality of the sensory signal is low (Peelle, 2018). Perceptual demands occur predominantly when audibility is poor or when competing stimuli partially mask the target. For example, natural listening environments may include speech masking introduced in the form of background noise, competing sounds (e.g., competing talkers), or degraded speech (e.g. speech heard over a public-announcement system). Speech masking is primarily
divided into two categories: 1) energetic masking and 2) informational masking. “Energetic masking” is the term used when two sounds interfere with each other on the basilar membrane of the inner ear. Two sounds (target and masker) physically mix and overlap in spectral content, so that one physically occludes the other. Energetic masking is typically contrasted with “informational masking”. If a signal is informationally masked, it is theoretically audible — the physical signal is not occluded at the periphery — but perception suffers for some other reason, like it is difficult to perceptually segregate from the masker because they are very similar in content. An example of informational masking occurs when speech from one talker is masked by speech from another (Kidd et al., 2007).

Processing demands that arise due to a signal being energetically masked differ from those arising due to informational masking. Degraded speech (i.e., energetic masking) recruits cognitive processes such as those engaged to use semantic context to help fill in missing signal as a means to increase intelligibility (Johnsrude and Rodd, 2016). In contrast, informational masking often occurs it is thought because the target and masker make similar processing demands on similar cognitive systems.

1.2.2 Linguistic demands
In addition to perceptual challenges, processing load can increase due to linguistic challenges. For example, both complex syntax (Walsh and Smith, 2011) and lexical ambiguity (Rodd et al., 2005) are cognitively demanding. Syntactic complexity of sentences increase speech-reception thresholds, with cognition explaining 40% of the variance (via PCA), suggesting that processing of syntactic complexity relies on individual cognitive capacities (Uslar et al., 2013). Processing of syntactic complexity also results in poorer accuracy and increased response latency (Walsh and Smith, 2011).

Wendt and colleagues (2016) were interested in comparing an acoustic and a linguistic challenge directly, examining their effects on pupil dilation. They tested Danish-speaking listeners and introduced a linguistic challenge by varying
sentence complexity. An idea in the form of a sentence was either presented in a more straightforward subject-verb-object structure (for example, "the angry penguin will film the sweet koala") or as a more syntactically complex object-verb-subject order, which is also correct in Danish, (for example, “the sweet koala, the angry penguin will film”). Both types of sentences were presented clearly, and also at a signal to noise ratio (SNR) of -6 dB, at which SNR the sentences are still mostly intelligible. The noise was created by overlapping 30 speech tracks. Thus, it had a rather flat envelope, and also had the long-term frequency spectrum of the sentence material.

Wendt and colleagues (2016) found that complex sentences resulted in more pupil dilation compared to simple sentences, and that background noise affected subjective ratings of effort (listeners report increased effort at the -6 dB SNR than in +12 dB SNR). In that study, a main effect of noise, but not complexity was found in the first epoch, whereas epochs 2 and 3 showed effects of complexity, but no noise. Wendt et al. showed an interaction of noise and complexity in epoch 3, yet they interpreted the effect as small and concluded that there were separable effects of noise level and complex syntax on pupil dilations and subjective effort. Subjective effort ratings are a common measure of cognitive load (see section 2.4 below). Syntactic complexity is one type of linguistic challenge; however, a more common linguistic challenge is semantic ambiguity, which is the one I investigate here.

1.3 Semantic Ambiguity
Many words in English have more than one meaning, with up to 80% of the words in an English dictionary having more than one word listed (Rodd et al., 2002). Sometimes these are different senses (both your ankle and the truth can be ‘twisted’ in somewhat different ways), but sometimes they are truly different meanings (e.g., bank of a river vs financial institution). Sometimes the same word can even have opposite meanings (e.g., sanction). Sentences containing words with more than one meaning, such as “the shell was fired towards the tank” pose
a particular challenge for cognition. Interpreting sentences containing ambiguous words can be effortful, this is especially important when the disambiguating context comes after the ambiguous word, and particularly when the correct meaning is not the dominant one. Disambiguation in such cases requires 1) maintenance of information in working memory, 2) inhibiting a prepotent interpretation, and then 3) selecting the correct interpretation based on all the words presented (Rodd et al., 2012). Although individuals are often unaware of the presence of ambiguity in natural language, behavioural and imaging studies demonstrate that such sentences increase processing load (Davis and Johnsrude, 2007; Rodd et al., 2012, 2010, 2005).

Processing of ambiguous sentences increases processing load, as is indicated by slower reaction times on a concurrent task (Rodd et al., 2010). Specifically, listeners were slower to judge whether a visually presented letter (e.g., “a”) was in upper or lower case, while they were simultaneously listening to sentences with ambiguous words (such as “There were dates and pears in the fruit bowl”), compared to psycholinguistically matched sentences without ambiguity (such as “There was beer and cider on the kitchen shelf”). Case judgments were significantly slower during high-ambiguity sentences, relative to low-ambiguity sentences. The increase in reaction time suggests an increased allocation of cognitive resources towards comprehending a high ambiguity sentence thereby leaving less cognitive reserve for case discrimination. This study suggests an overlap between the cognitive system involved in semantic disambiguation and the domain-general process of response selection required for the case-judgement task. This cognitive overlap may reflect neural overlap in the networks supporting these processes and is consistent with the proposal that domain-general selection processes in inferior frontal regions are critical for language comprehension (Rodd et al., 2010).
1.4 Neuroimaging correlates of semantic ambiguity

Functional magnetic resonance imaging (fMRI) evidence indicates that different types of listening challenges recruit anatomically distinguishable brain networks. For example, degrading sentences acoustically by adding noise recruits the left lateral temporal cortex, inferior frontal cortex, and posterior middle frontal gyrus (Davis and Johnsrude, 2003), with decreasing intelligibility increasing cingulo-opercular and fronto-parietal cortex (Davis and Johnsrude, 2003; Golestani et al., 2013; Vaden et al., 2013) activity. In contrast, hearing sentences with ambiguous words, compared to low-ambiguity control sentences (i.e., the materials described in the previous section) generates activity in left inferior frontal gyrus (Rodd et al., 2012, 2005), and posterior inferior temporal lobe regions.

An unpublished fMRI study from our laboratory examined brain activity in response to sentences with and without ambiguous words (the same high- and low-ambiguity sentences compared above), presented either clearly or in background noise, in a factorial design (manuscript in preparation). The two challenges resulted in very different patterns of activity, as shown in Figure 1A, with two regions – in the left anterior insula and anterior cingulate -- exhibiting overlap between the two challenges (Fig. 1A), and a small dorsal inferior frontal gyrus region exhibiting an interaction effect. The anterior insula and cingulate regions are consistent with the cinguloopercular network, which has repeatedly been demonstrated to be active during acoustic challenges to speech comprehension (Eckert et al., 2016; Peelle, 2018; Vaden et al., 2013). This suggests that the semantic challenge of ambiguous words recruits brain networks that are largely – although not entirely - independent of those recruited for the acoustic challenge of masking noise. This in turn indicates that not all challenges to comprehension rely on the same cognitive resources for compensation. According to the self-reported effort in both young and old participants in this study, linguistic ambiguity is further hampered by noise, such that high ambiguity sentences in conjunction with noise produce a further decrease in intelligibility, relative to clear conditions. That is,
increased cognitive load resulting from ambiguity leaves less cognitive load for processing acoustics.

Rodd et al (2005) showed extensive differential brain network activation to high-vs. low-ambiguity sentences (Fig. 1B). Participants had to listen to these sentences, and at the end of each sentence, they indicated if a word presented visually was related to that sentence or not (i.e., a sentence comprehension task to make sure participants processed meaning). The network associated with speech comprehension, independent of ambiguity, showed bilateral activation in the superior and middle temporal gyri. Comparisons of high vs. low ambiguity sentences showed greater activation in the left and right inferior frontal gyrus (left > right), in addition to left temporal region which includes posterior inferior temporal cortex, middle temporal gyrus and superior temporal sulcus (Fig. 1B; yellow and red blobs). These areas may participate in cognitive processing of disambiguating information. Interestingly, participants in this study underestimated the amount of ambiguity present, despite the increased brain activity (i.e., increased brain effort). Thus, self-reported data on cognitive effort may not capture the demand imposed on the brain by ambiguous words, and perhaps a non-volitional or autonomic indicator of processing is needed to observe effort objectively.
Figure 1. Neuroimaging correlates of semantic ambiguity and noise. (A) Activation image showing main effects of Clear – Noise (in blue), of High-ambiguity – Low-ambiguity (in red) and the interaction between Noise and Ambiguity [(HAclear-LAclear)-(HAnoise-LAnoise)] (in green), superimposed on the average structural for the group. Yellow indicates regions exhibiting both main effects to a significant degree; cyan indicates regions in which both a main effect of Ambiguity and a Noise by Ambiguity interaction are present. These overlapping regions are in the anterior insula and anterior cingulate, consistent with the cinguloopercular network. (B) from Rodd et al., 2005: High-ambiguity sentences resulted in greater activity relative to low-ambiguity sentences in the inferior frontal sulcus (bilaterally but left > right) and the left temporal region which includes posterior inferior temporal cortex, middle temporal gyrus and superior temporal sulcus (yellow and red).
1.5 Pupillometry

In clinical audiology research, there is a growing interest in measuring pupil size (pupillometry) as a way to objectively characterize effortful listening, which has typically been assessed using subjective questionnaires (Kuchinsky et al., 2014; McGarrigle et al., 2017). Pupillometry is cost-effective and user-friendly, making it ideal for clinical practice.

A host of studies indicate that when acoustics are degraded or if hearing impairment is present then the pupil is larger, compared to when listening is effortless (Zekveld et al., 2011; Koelewijn et al., 2012; Ohlenforst et al., 2017; Zekveld and Kramer, 2014). Larger pupil size is observed for: higher levels of degradation (McGarrigle et al., 2017; Miles et al., 2017), different types of degradation (i.e., informational masker > energetic masker) (Koelewijn et al., 2012; Ohlenforst et al., 2017a), during poorer performance (Miles et al., 2017), and when sound level is increased (Nunnally et al., 1967).

Fifty years of pupillometry research highlights that after controlling for luminance and accommodation, any changes in pupil size also reflect arousal and mental effort (Kahneman and Beatty, 1966). A plethora of cognitive work linking pupil dilation with attention (Laeng et al., 2012), memory (Johnson, 1971), processing load (Beatty, 1982), and executive functions (van der Wel and van Steenbergen, 2018) indicates that pupil dilation increases with increasing task demands. According to Kahneman and Beatty, “anything that increases the brain’s processing load, will dilate the pupil” (Kahneman and Beatty, 1966). If linguistic and acoustic factors both influence processing load during speech, then the pupil should reflect both these factors. Indeed, recently it has been shown that syntactic complexity and energetic masking using a stationary speech shaped noise both result in pupil dilation (Wendt et al., 2016).

Exactly what processes lead to pupil dilation is not clear. Cognitive challenge may lead to an increase in arousal, and that may be what pupil dilation is reflecting directly.
1.5.1 Physiology of the pupillary response

High-level cognition (via increased arousal) is not the only stimulus that can cause pupil dilation (Mathôt, 2018). The primary function of the pupil is to regulate light entry into the eye (luminance-dependent modulation) and help focus an image (accommodation). The onset of any salient stimulus, whether visual, auditory, or other, can induce pupil dilation (Wang and Munoz, 2015).

Pupil size is dependent on the interplay of two branches of the autonomic nervous system: the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The size of the pupil is controlled by two muscles that are located in the iris (the colored area around the pupil). The iris sphincter muscle, innervated by the PNS (homeostasis, rest-and-digest) causes the pupil to constrict, and the iris dilator muscle, innervated by the SNS (i.e., arousal, fight-or-flight response) causes the pupil to dilate (Mathôt, 2018). The net pupil size at any given moment is a balance between the opposing effects of these two systems.

The primary neuronal pathway that underlies pupil dilation (i.e., sympathetic dilation pathway) begins at the at the hypothalamus and the locus coeruleus (LC). Both these areas are involved in arousal, have reciprocal excitatory connections and project to the intermedio-lateral column of the spinal cord. Post-synaptic neurons travel down through the brain stem and exit through the cervical sympathetic chain and the superior cervical ganglion to the iris dilator muscle (a group of muscles in the peripheral 2/3 of the iris). The pathway of pupillary constriction (i.e., parasympathetic constriction pathway) begins at the Edinger-Westphal nucleus (near the oculomotor nerve nucleus). The fibers enter the orbit with nerve fibers of the oculomotor cranial nerve and ultimately synapse at the ciliary ganglion, then to the circular iris sphincter muscle. Although the pathways are distinct, there are some overlapping neural interactions: the locus coeruleus causes pupil dilation not only by activating the SNS but also by inhibiting PNS via Edinger-Westphal nucleus inhibition (Steinhauer et al., 2004).
Pupil size is a reliable index of LC activity. Electrical stimulation of the locus coeruleus in mice, rats, and nonhuman primates precedes, and subsequently correlates, with large and rapid pupil dilations (Joshi et al., 2016; Liu et al., 2017; Reimer et al., 2016). The pupil also changes during pharmacological manipulations of arousal (via modulation of LC) (Hou et al., 2005). Furthermore, fMRI studies have linked regional LC activation to pupil changes (Murphy et al., 2014). The LC is the main source of norepinephrine production in the brain. It is located in the dorsolateral pontine of the brainstem and innervates much of the neocortex, including the fronto-parietal network associated with executive functions (e.g., working memory, goal-directed behaviour). The locus coeruleus-norepinephrine system also has a major role in attention and arousal (Schneider et al., 2016). The pathway that involves inhibition of the Edinger-Westphal Nucleus via LC has been suggested to underly pupil dilation in response to arousal and mental effort (Steinhauer et al., 2004).

1.6 Microsaccades

Saccades are fast eye movements that shift sight from one target to another (Wang et al., 2017). Microsaccades are small, involuntary saccadic eye movements that occur 1-2 times per second during prolonged visual fixation (Costela et al., 2013; Privitera et al., 2014). They may have an impact on pupil dilations and are important to consider for pupillary activity studies (Knapen et al., 2016). The neural circuitry underlying pupil dilation and microsaccade rate are linked, as the respective anatomical domains, the locus coeruleus and the superior colliculus, are neurally connected. The intermediate layers of the superior colliculus innervates pupil circuitry as it receives input from the locus coeruleus (Edwards et al., 1979), projects to the Edinger-Westphal Nucleus (Harting et al., 1980), and the pupil dilates upon stimulation of the superior colliculus (Wang et al., 2012). Moreover, pupil dilations and microsaccade rate are correlated with each other (Raj et al., 2019). Interestingly, cognitive load is linked to changes in both pupil size and microsaccade rate, with cognitive load apparently decreasing the rate of
microsaccades (Gao et al., 2015). To this end, we considered the effect of microsaccade rate when disseminating our findings.

1.7 Rationale for the current study
Acoustic degradation results in pupil dilation, which is thought to reflect listening effort. Syntactic complexity also results in pupil dilation, presumably because of a more abstract kind of mental effort, not directly attributable to listening. Semantic ambiguity may also dilate the pupil, which would demonstrate that pupillometry can be sensitive to multiple challenges involved with speech reception. This would further indicate that pupil dilation is a nonspecific marker of difficulty, and that the linguistic challenge(s) introduced by diverse experimental materials needs to be carefully controlled if pupil response to an acoustic challenge is to be interpretable. For example, pupil measures in audiology may hinder the interpretation of a clinical population as individual differences in the cognitive abilities recruited to cope with ambiguous words may manifest as differences in pupil dilation, independent of hearing ability.

Here, we use pupil dilation and microsaccade measures to investigate the impact of semantic ambiguity and noise in conditions that modulate both the acoustic and linguistic dimensions of effort. Specifically, using a factorial design, we delivered low and high ambiguity sentences in both clear and noise conditions. We aim to investigate how effort under linguistic (i.e., semantic ambiguity) demands and perceptual (i.e., noise) is reflected in the pupillary response.

In Experiment 1, we investigate the effect of semantic ambiguity and an energetic masker (overlap in spectral content) on pupil size. In Experiment 2 we wanted to be able to compare our results to those of Wendt et al (2016), using semantic ambiguity instead of syntactic complexity. The masker used by Wendt et al, and which we use in Experiment 2, is a babble noise, which has elements of both energetic and informational masking (perceptual segregation). Linguistic (e.g., phonetic) information in the babble noise that perhaps engages resources that are
also needed to process the target (i.e., the ‘informational masking’ component for this experiment). Given that the babble noise envelope is flat (much flatter than the envelope in Experiment 1) and minimally fluctuating, there may be windows of improved hearing (glimpsing) of the target sentence, when the amplitude of the target is considerably higher than the babble noise signal (Bologna et al., 2018). This is in contrast to the constant noise used in Experiment 1, which had exactly the same envelope as the masked sentence and so fluctuates with the target sentence, providing absolutely constant SNR. This noise does not provide a window where the target sentence has a greater amplitude than the noise. Wendt et al (2016) also had the noise start earlier than the sentence, perhaps engaging segregation processes that would not be engaged in Experiment 1. Thus, by using a different noise, and starting it earlier than the target, we think we are engaging a different set of cognitive processes during listening. Do the effects on the pupil evident in experiment 1 replicate, under these different listening conditions?

The objectives of the current study are: (1) to characterize differential pupillary, and microsaccadic responses to semantic ambiguity and background noise and (2) to explore the impact of different types of background noise on pupillary responses. The overall hypothesis is that the presence of a masker and the presence of ambiguous words when listening to speech both lead to increases in processing load and increased pupil dilation. We predict that both ambiguity and noise will influence the pupillary response, and if these are independent effects that both act on the pupil then their effect will be additive.
Chapter 2

2. Methods

2.1 Participants
Seventy-three graduate and undergraduate students from The University of Western Ontario (Canada) were recruited in two experiments (N=35 for Experiment 1, mean age: 19 years, range: 18-21 years, 15 females; N=38 for Experiment 2, mean age: 19 years, range: 18-33 years, 23 females). Data from one additional participant were excluded due to failure in data storage (N=1; Exp 1). Participants self-reported having normal hearing, normal or corrected to normal vision, and no neurological disorders in their history. Participants gave written informed consent and received course credits or were paid $10 per hour for their participation. The experimental protocols were approved by the Research Ethics Board of the University of Western Ontario (protocol ID: HSREB 106570; See Appendix B) and are in line with the declaration of Helsinki.

2.2 Auditory stimuli and task
We utilized sentence materials from previous studies, in which the effect of sentence ambiguity on behaviour and on brain activity were investigated (Rodd et al., 2005, 2010). In the high-ambiguity condition, sentences contained two or more ambiguous words (e.g., The shell was fired towards the tank). In the low-ambiguity condition, sentences contained no ambiguous words (e.g., Her secrets were written in her diary). Original sentences were in British English and they were re-recorded in Canadian English. The duration of sentences ranged from 1.4 s to 4.8 s. The speech stimuli in the low-ambiguity and high-ambiguity conditions were matched on duration and psycholinguistic parameters (words, imageability, concreteness, and word frequency (Rodd et al., 2005). A 2 x 2 factorial within-subject design (sentence ambiguity x background noise) was used to present 56 low- and 56 high-ambiguity sentences. Sentences in the
four conditions were presented in four blocks. Seven sentences per condition were presented within each block (N=28 trials per block), for a total of 112 sentences in each experiment. Sentences were presented pseudo-randomly such that no more than three sentences of the same ambiguity condition and two sentences of the same noise condition could occur in a row. Half of the low- and high-ambiguity sentences were presented under clear conditions, the other half were presented in background noise. Background noise conditions (clear/low noise, high noise) were counterbalanced across participants. Each participant heard each stimulus only once.

In Experiment 1 (Figure 2A), sentences were either presented under clear conditions (no noise) or with added background noise (noise). The background noise was created uniquely for each sentence by modulating pink noise (each octave carries equal power) by the sentence’s amplitude envelope. In detail, the sentence envelope was extracted by calculating the Hilbert transform of the original sentence, followed by low-pass filtering (30-Hz Butterworth) of the absolute value of the complex numbers resulting from the Hilbert transform. The envelope was used to modulate pink noise. The original sentence and the sentence-specific modulated pink noise were added at a signal-to-noise ratio of -2 SNR (i.e., the sentence was 2 dB lower in level relative to the modulated pink noise). Pink noise is a broad-spectrum noise with 1/f structure that provides energetic masking. Since the signal and masker have the same envelope, the masking level (degree of energetic masking) is constant over the period. All stimuli (including clear and those with noise added) were matched in their root-mean-square intensity level.

For Experiment 2 (Figure 2B), high- and low-ambiguity sentences were presented either under low (6 SNR) or under high background noise (0 SNR), in a factorial design with materials counterbalanced across participants, as for Experiment 1. Background noise was a 30-talker babble with a long-term frequency spectrum of the current sentence materials. The noise was created by adding 30 tracks of randomly concatenated sentences (Wagener et al., 2003). The babble noise
envelope is more stationary (i.e., less modulated) compared to the noise envelope in Experiment 1, resulting in temporal glimpses of the target speech from the background noise (Bologna et al., 2018). This is in contrast to the pink noise which fluctuates with the target sentence and does not provide a window where the target sentence could have a greater amplitude than the noise. Moreover, there is linguistic (e.g., phonetic) information in the babble noise that perhaps engages resources that are also needed to process the target (i.e., the ‘informational masking’ component for this experiment). The original sentence was added to the background noise such that the noise started three seconds before sentence onset and ended 1.2 s after sentence offset. The noise level was kept constant for all conditions (which avoids providing cues as to which condition is presented), while the level of the sentence was adjusted to a signal-to-noise ratio of 6 dB (low SNR) or to a signal-to-noise-ratio of 0 dB (high SNR) depending on the condition.

Figure 2. Experimental design. A: Trial schematic of experiment 1. B: Trial schematic of experiment 2.

2.3 Procedure and data recording
Participants were tested in a dim, quiet room while wearing headphones (Sennheiser HD 25-SP II). Sentences were presented via a Steinberg UR22 (Steinberg Media Technologies) external sound card. Experimental procedures were controlled using Psychtoolbox in MATLAB (v2015b, Mathworks Inc.). Prior to
the main experimental procedures, the sensation level was determined for each participant using a method-of-limits procedure (Herrmann and Johnsrude, 2018). This procedure entailed alternating trials of progressively increasing or decreasing 12-second long tones over time by 5.4 dB/s. Participants indicated when they could no longer hear the tone (progressively decreasing intensity trial) or when they started to hear the tone (progressively increasing intensity trial). Each of the progressively increasing and decreasing intensity trials were conducted six times, and at the time of the button press, the corresponding mean sound intensity during the trial was collected. Finally, the six trials for each trial type were averaged to determine the individual hearing threshold. In both experiments, sounds were presented at 45 dB above the individual's sensation level.

During the experiments, participants rested their head on a chin and forehead rest (EyeLink 1000 Tower mount) facing a screen at a distance of 670 mm. Pupil area and eye movements were recorded continuously from the left eye using an integrated infrared camera (eye tracker 1000; SMI, Needham, MA) at a sampling rate of 500 Hz. Nine-point fixation was used for eye-tracker calibration (McIntire et al., 2014).

During the experiments, each trial was structured as follows. A fixation circle (white on grey background) was present, starting three seconds before sentence onset, and remained on screen until 1.2 s after sentence offset. In Experiment 1, no sound stimulation was administered during the three seconds prior to sentence onset and the 1.2-s post-sentence offset. In Experiment 2, the speech-shaped noise was presented throughout that period, that is, from three seconds prior to sentence onset until 1.2 s post-sentence offset (Figure 2B). A sentence was played three seconds after the onset of the fixation circle. The fixation circle on the screen was replaced by a semantic relatedness task 1.2 s after sentence offset. A probe word was presented on the screen (e.g., “Book”), and participants had to indicate with a keypress whether or not this word was semantically related to the sentence they had heard. The word remained on screen for 3.5 seconds or until participants
pressed the related (left index finger) or unrelated (right index finger) button on a keyboard, whichever came first. The semantic relatedness task was chosen as a way to ensure that participants were attending to the sentences. Finally, the screen was cleared, between trials, for 5-7 seconds in order to allow participants to rest and blink.

Participants underwent a training block of 8 trials (using sentences not used in the experiment) before the experiment in order to familiarize them with the experimental procedures (including eye-tracker calibration). The experiment took approximately one hour to complete.

2.4 Data analysis

Data analysis was carried out offline using custom MATLAB scripts (v2018b), and the analyses were identical for both experiments.

2.4.1 Behavioural analysis

Behavioural data of the semantic ambiguity task were analyzed by calculating the mean number of correct responses, separately for each ambiguity and noise condition. A correct response entailed responding with “related” when a word was semantically related to the preceding sentence or by pressing “unrelated” when the word was not semantically related to the preceding sentence. Separately for each experiment, a 2 × 2 repeated-measures analysis of variance (rmANOVA) was calculated, with factors Semantic Ambiguity (low vs. high) and Noise (Exp 1: clear vs. noise; Exp 2: high vs. low SNR). A significant interaction was resolved by subsequent t-tests.

2.4.2 Pupillometry analysis

Preprocessing of pupil area involved removing eye blink artifacts. For each eye blink indicated by the eye tracker, all data points between 50 ms before and 200 ms after a blink were removed. In addition, pupil area values that differed from the median pupil area by more than three times the median absolute deviation (MAD) were classified as outliers and removed (Leys et al., 2013). Missing data resulting
from artifact rejections and outlier removal were linearly interpolated. Data for an entire trial were excluded from analysis if missing data made up more than 40% of the trial (Exp 1, N=63/3920; Exp 2, N=72/4256). Data were low-pass filtered at 10 Hz (Kaiser window, length: 201 points). Single-trial time courses were baseline-corrected by subtracting the mean pupil size from the -0.5 s to 0 s time window from each pupil size value (Mathôt et al., 2018). Single-trial time courses were averaged separately for each condition, and displayed for the -0.5 s to 4 s epoch.

Three dependent measures were extracted: Mean pupil dilation, peak pupil dilation, and peak pupil latency (Winn et al., 2018). Critically, in order to account for the different sentence durations at the analysis stage, mean pupil dilation was calculated for each trial as the average pupil area for the time window ranging from 0.5 s after sentence onset to 1 second after sentence offset, and subsequently averaged across trials, separately for each condition and participant. Peak and latency to peak pupil dilation were extracted for each trial within the time window ranging from 0.5 s after sentence onset to 1 second after sentence offset, and subsequently averaged across trials, separately for each condition and participant.

Separately for each experiment and each dependent measure, a $2 \times 2$ rmANOVA was calculated, with factors Semantic Ambiguity (low vs. high) and Background noise (Exp 1: clear vs. noise; Exp 2: high vs. low SNR). A significant interaction was resolved by subsequent t-tests.

### 2.4.3 Microsaccade analysis

Despite instructions to maintain fixation and reduce blinks, variability in pupil responses may stem from ocular events (e.g., microsaccades). Microsaccades occur during prolonged fixation (Widmann et al., 2014), as was used here, and may influence pupil dilation (Knapen et al., 2016). We therefore tested the extent to which microsaccades may show effects due to our experimental manipulations. Microsaccades were identified using a method that computes thresholds based on velocity statistics from eyetracker data, and then identifies outliers (Engbert, 2006; Engbert and Kliegl, 2003). That is, the vertical and horizontal eye movement time
series were transformed into velocities and microsaccades were classified as outliers that exceed: (1) a relative velocity threshold of 15 times the median-based standard deviation of the eye-movement velocity and (2) where the conditions of 1) persisted for 6 ms or longer (Engbert and Kliegl, 2003). A time course of microsaccade rate was calculated from the individual microsaccade times (Widmann et al. 2014) by convolving each microsaccade occurrence with a Gaussian window (0.02 ms). For analysis purposes, mean microsaccade rate was calculated across trials as the average rate in the time window 0.5 s before to 1 s after sentences offset, then subsequently averaged across trials. For displaying purposes, mean microsaccade rate was calculated across trials as the average rate in the time window 0.5 to 4 s time locked to sentence onset. Mean microsaccade rate was subsequently averaged across trials.

Separately for each experiment, a $2 \times 2$ rmANOVA was calculated for the mean microsaccade rate, with factors Semantic Ambiguity (low vs. high) and Background noise (Exp 1: clear vs. noise; Exp 2: high vs. low SNR). A significant interaction was resolved by subsequent t-tests.
3 Results

3.1 Experiment 1

3.1.1 Semantic relatedness task
Figure 3 shows a bar graph for each condition in the semantic relatedness task. The rmANOVA for the correct responses revealed a main effect of Background Noise ($F_{1,34} = 22.69, p = 3.467 \times 10^{-5}, \eta^2_p = 0.4$), showing lower performance when sentences were presented in noise compared to clear sentences. The effect of Semantic Ambiguity was not significant ($F_{1,34} = 0.879, p = 0.354, \eta^2_p = 0.025$). The Semantic Ambiguity $\times$ Background Noise interaction was significant ($F_{1,34} = 6.23, p = 0.017, \eta^2_p = 0.155$). Performance was lower for high ambiguity compared to low-ambiguity sentences in noise ($t_{34} = 2.369, p = 0.023$), but not in clear ($t_{34} = 1.74, p = 0.089$).

![Figure 3. Semantic relatedness results for Experiment 1. Bar graph shows mean correct responses for each condition. Error bars reflect the standard error of the mean. LAC – low ambiguity in clear, HAC – high ambiguity in clear, LAN – Low ambiguity in -2SNR noise, HAN – High ambiguity in -2SNR noise.]

3.1.2 Pupillometry
Figure 4A shows the time course for the pupil area. The rmANOVA for the mean pupil area revealed a main effect of Background Noise ($F_{1,34} = 55.68, p = 1.169 \times 10^{-8}, \eta^2_p = 0.621$), showing that the pupil area was greater in the noise conditions.
compared to the clear conditions (Figure 4B,F). The main effect of Semantic Ambiguity was also significant \( (F_{1,34} = 5.53, p = 0.024, \eta^2_p = 0.14) \), showing that the pupil area was greater in the high-ambiguity compared to the low-ambiguity condition (Figure 4B,E). The Semantic Ambiguity × Background Noise interaction was not significant \( (F_{1,34} = 1.80, p = 0.188) \). Given that the time courses depicted in Figure 4A suggest a smaller effect of semantic ambiguity under noise compared to clear conditions, we explored this possibility. An analysis of simple effects indeed showed that the effect of ambiguity was significant for the clear \( (t_{34} = 2.16, p = 0.037) \) but not for noise conditions \( (t_{34} = 0.793, p = 0.432) \).

The results for peak pupil area mirrored those from the mean pupil area (Figure 4C). A main effect of Background Noise \( (F_{1,34} = 53.57, p = 1.77^{-8}, \eta^2_p = 0.611) \) and a main effect of Semantic Ambiguity \( (F_{1,34} = 6.729, p = 0.0139, \eta^2_p = 0.165) \) were observed. Peak pupil dilation was larger in the noise compared to the clear condition, and greater in the high-ambiguity compared to the low-ambiguity condition.
condition. The Semantic Ambiguity $\times$ Background Noise interaction was not significant ($F_{1,34} = 0.2834, p = 0.283$).

The results for peak latency (Figure 4D) revealed a main effect of Semantic Ambiguity ($F_{1,34} = 11.48, p = 0.001$), but not for Background Noise ($F_{1,34} = 0.221, P = 0.640$). Pupil size peaked later in the high-ambiguity compared to the low-ambiguity condition. The Semantic Ambiguity $\times$ Background Noise interaction was not significant ($F_{1,34} = 1.8029, p = 0.188$).

In sum, the results from Experiment 1 show that pupil dilation is sensitive to background noise and semantic ambiguity, showing that acoustic and linguistic factors affect pupil dilation. In Experiment 1, we used a modulated pink noise as a masker that predominantly masks the speech signal energetically. In Experiment 2, speech-shaped noise was used as a masker in order to investigate whether pupil dilation is also sensitive to acoustic and linguistic factors under conditions that resemble more closely challenging listening situations of everyday life in which informational and energetic masking occur concurrently, similar to a crowded restaurant.

### 3.2 Experiment 2

#### 3.2.1 Semantic relatedness task

Figure 5 shows the results for the semantic relatedness task. The rmANOVA for the correct responses revealed a main effect of Background Noise ($F_{1,37} = 45.133, p = 6.74^{-8}, \eta^2_p = 0.549$), with fewer correct responses in the low SNR (0 dB SNR) condition compared to the high SNR (6 dB SNR) condition. The main effect of Semantic Ambiguity was not significant ($F_{1,37} = 3.148, p = 0.084, \eta^2_p = 0.078$). The Semantic Ambiguity $\times$ Background Noise interaction was significant ($F_{1,37} = 8.118, p = 0.007, \eta^2_p = 0.179$). In the low SNR condition, high-ambiguity sentences resulted in fewer correct responses compared to low-ambiguity sentences ($t_{37} = 3.03, p = 0.004$), whereas this contrast was not significant for the high SNR condition ($t_{37} = -0.544, p = 0.589; 93\% - 91\%$).
Figure 5. Semantic relatedness results for Experiment 2. Bar graph shows mean correct responses for each condition. Error bars reflect the standard error of the mean. LA6 – low ambiguity in 6 dB SNR noise, HA6 – high ambiguity in 6 dB SNR noise, LA0 – low ambiguity in 0 dB SNR noise, HA0 – high ambiguity in 0 dB SNR noise.

3.2.2 Pupillometry

Pupil area time courses are displayed in Figure 6A. Similar to Experiment 1, the rmANOVA on mean pupil area revealed a mean effect of Background Noise ($F_{1,37} = 10.3$, $p = 0.002$, $\eta^2_p = 0.218$), showing that the pupil area was larger in the low SNR condition (0 dB SNR) compared to the high SNR condition (6 dB SNR; Figure 6B,F). In addition, the main effect of Semantic Ambiguity was marginally significant ($F_{1,37} = 3.73$, $p = 0.061$, $\eta^2_p = 0.091$; Figure 6B,E), indicating that the pupil area tended to be greater in the high-ambiguity compared to the low-ambiguity condition. The Background Noise × Semantic Ambiguity interaction approached significance ($F_{1,37} = 3.90$, $p = 0.055$, $\eta^2_p = 0.095$). In order to explore the trending interaction, we analyzed the simple effects. Pupil area was larger in high-ambiguity compared to low-ambiguity sentences under high SNR (i.e., low noise; $t_{37} = 2.953$, $p = 0.0054$), but not under low SNR (high-noise; $t_{37} = 0.3552$, $p = 0.724$). Pupil area was also larger for low-SNR (high-noise) sentences compared to noise was significant for the low ambiguity ($T_{37} = 3.32$, $p = 0.002$) but not for the high ambiguity ($T_{37} = 0.681$, $p = 0.5$) condition.

The results for peak pupil area revealed a main effect of Background Noise ($F_{1,37} = 18.1$, $p = 1.3e^{-4}$, $\eta^2_p = 0.328$) and a main effect of Semantic Ambiguity ($F_{1,37} = 4.7212$, $p = 0.036$, $\eta^2_p = 0.113$). Peak pupil dilation was greater in the noise compared to the clear condition, and greater in the high- compared to the low-
ambiguity condition. The Semantic Ambiguity × Background Noise interaction was not significant ($F_{1,37} = 2.19, p = 0.1465$).

The results for peak latency revealed no significant main effects (Background Noise: $F_{1,37} = 0.26, p = 0.61$; Semantic Ambiguity: $F_{1,37} = 3.48, p = 0.069$) and no interaction ($F_{1,37} = 0.53, p = 0.468$; Figure 6D).

Figure 6. Pupil dilation results for Experiment 2. A: Time course of pupil area (averaged across participants; N=38) for sentences between 2 and 3 s in duration (74/112). Histogram below shows distribution of sentence duration. B: Mean pupil area from 0.5 s after sentence onset to one second after sentence offset. C: Peak pupil dilation. D: Latency of peak pupil dilation. Error bars reflect 95% confidence intervals. E: Individual data scatter plot for Semantic Ambiguity main effect (N=38); pupil mean area. F: Individual scatter plot for Background Noise main effect (N=38). LA6 – low ambiguity in 6 dB SNR noise, HA6 – high ambiguity in 6 dB SNR noise, LA0 – low ambiguity in 0 db SNR noise, HA0 – high ambiguity in 0 dB SNR noise.

### 3.3 Pooling Experiment 1 and 2

In order to gain more statistical power for the analysis of the effect of semantic ambiguity on pupil dilation, we pooled the data from Experiment 1 and 2, leading to 73 participants entered into the analysis. A rmANOVA on mean pupil dilation was analyzed as before, with Experiment as a between-subjects factor. The rmANOVA revealed a main effect of Semantic Ambiguity ($F_{1,71} = 9.32, p = 0.003, \eta_p^2 = 0.116$), of Background Noise ($F_{1,71} = 69.7, p = 3.7e^{-12}, \eta_p^2 = 0.496$), and a Semantic Ambiguity × Background Noise interaction ($F_{1,71} = 4.97, p = .029, \eta_p^2 = \ldots$)
Simple effects revealed that pupil area was larger in high-ambiguity compared to low-ambiguity sentences under high SNR (i.e., low noise; $t_{72} = 3.276$, $p = 0.002$), but not under low SNR (high-noise; $t_{72} = 0.351$, $p = 0.726$). A rmANOVA on microsaccade rate was analyzed, with Experiment as a between-subjects factor. The rmANOVA revealed a main effect of Experiment ($F_{1,71} = 12.08$, $p = 0.001$, $\eta^2_p = 0.145$), indicating lower microsaccade rate in Experiment 2, relative to Experiment 1.

### 3.4 Microsaccade results

Microsaccades were analyzed in order to investigate whether saccadic eye movements during fixation are also sensitive background noise and semantic ambiguity. Microsaccade time courses are depicted in Figure 7. Microsaccade rate averaged across 0.5 s post-sentence onset to 1 s post-sentence offset are shown in Figure 7C and D. No significant main effects or interactions were observed in the two experiments (Experiment 1: Background Noise: $F_{1,34} = 0.039$, $p = 0.844$, $\eta^2_p = 0.001$; Semantic Ambiguity: $F_{1,34} = 0.122$, $p = 0.729$, $\eta^2_p = 0.003$; Background Noise × Semantic Ambiguity interaction: $F_{1,34} = 0.017$, $p = 0.895$, $\eta^2_p = 5\times10^{-4}$; Experiment 2: Background Noise: $F_{1,37} = 0.051$, $p = 0.821$, $\eta^2_p = 0.001$; Semantic
Ambiguity: $F_{1,37} = 0.003$, $p = 0.956$, $\eta^2_p = 8 \times 10^{-5}$; Background Noise $\times$ Semantic Ambiguity interaction: $F_{1,37} = 0.316$, $p = 0.577$, $\eta^2_p = 0.008$.

**Figure 7. Results for microsaccade analysis.** Time courses for microsaccade rate for Experiment 1 (A) and Experiment 2 (B). Bar graphs show the mean microsaccade rate for each condition for Experiment 1 (C) and Experiment 2 (D). Error bars reflect the standard error of the mean. MS – microsaccade.
4 Discussion

In two experiments, we investigated how listening to sentences with different permutations of noise and semantic ambiguity levels affected: 1) performance on a semantic relatedness task; and 2) pupil dilation; and 3) microsaccades. When participants listen to sentences containing ambiguous words, mean and peak pupil size increased relative to sentences with low ambiguity. Two different kinds of noise were used. These included a pink noise modulated with a target sentence’s envelope to yield a constant SNR (experiment 1), and a more complex multi-talker babble noise that had a flatter envelope and provided informational in additional to energetic masking. In addition to a well-established pupil dilation observed for both types of background noise, I observed an effect of semantic ambiguity.

4.1 Behavioural effects

The semantic-relatedness task tapped sentence comprehension, and so required listeners to pay attention to the meaning of a sentence. This behavioural task was necessary to ensure engagement and input of effort from the participant during listening such that there would be an effort-based stimulus to elicit pupil dilation (Hopstaken et al., 2015). In Experiments 1 and 2, there was a clear main effect of noise on performance, and an interaction between noise and ambiguity such that the effect of ambiguity on performance was not evident when sentences were clear (Experiment 1) or when noise level was low (Experiment 2). This may be a ceiling effect: the clear conditions of Experiment 1 probably define the upper limit on performance, and performance was highly consistent across both levels of ambiguity in the low noise conditions of Experiment 2 compared to the clear conditions of Experiment 1 (89 – 93 % across the four conditions). However, these levels are well below 100%, and an alternative explanation, consistent with imaging data (Rodd, 2005, 2012), is that the presence of ambiguity recruits additional cognitive processes that enhance comprehension. When perceptual
demand is low (i.e., clear condition in Experiment 1 and 6 dB SNR in Experiment 2), these processes compensate effectively, and comprehension is normal. When perceptual load is higher (the two higher noise conditions), these processes are not adequate, and comprehension suffers. This interpretation is also consistent with the pupillary effects we observed in both experiments, with a clear effect of ambiguity (reflecting cognitive load) evident regardless of noise level.

4.2 Pupillary effects
Experiments 1 and 2 showed increased mean pupil area and peak pupil area when sentences contained ambiguous words (compared to matched sentences without) and when sentences were masked by noise relative to clear speech (Experiment 1) or when the speech-to-noise level was reduced (Experiment 2). In Experiment sentences were masked by a pink noise with the same envelope as the sentence, yielding a constant SNR. In Experiment 2, sentences were masked by a multi-talker babble noise with a flatter envelope compared to the pink noise.

The rmANOVA suggested that these two factors are independently influencing pupil size, since no significant interaction was observed for mean or peak pupil dilation. However, pooling the data of Experiment 1 and 2 suggest that semantic ambiguity influenced pupil dilation less when noise was present or the SNR was lower than when sentences were presented clearly or at a higher SNR, respectively. Moreover, in Experiment 1, pupil dilation peaked later for high-ambiguity compared to for low-ambiguity sentences (Figure 4D). Together, these data show that pupil dilation occurs in response to linguistic (in the form of semantic ambiguity) as well as perceptual (i.e., noise) demands.

4.3 Microsaccade effects
The current and previous data show that pupil dilation is sensitive to acoustic and linguistic challenges during speech comprehension. However, pupil size is also sensitive to saccades and microsaccades (Knapen et al., 2016). Participants
fixated on a circle at the center of the screen and thus microsaccades could in principle be entangled with the changes in pupil size that we observed.

Here we investigated the effects of our conditions on microsaccade rate, but no effects were found. These results make it unlikely that microsaccadic eye movements confound the effects of acoustic and cognitive challenges on pupil dilation observed in the current work. In line with previous studies (Rolfs, Kliegl, & Engbert, 2008; Rolfs, Engbert, & Kliegl, 2005), we observed microsaccade inhibition at the onset of our auditory stimuli (Figure 7A). Microsaccade inhibition is characterized by the appearance of an initial dip in microsaccade rate, followed by an overshoot and a return to baseline at the onset of visual or auditory stimuli (Rolfs et al., 2008). Analysis of microsaccade differences (See Section 3.3) between experiments shows that the microsaccade rate was overall lower in Experiment 2 compared to Experiment 1 (Figure 7). Microsaccade rate has been shown to decrease with high cognitive load (Xue et al., 2017), working memory load (Dalmaso et al., 2017), and task difficulty (Siegenthaler et al., 2014). Whether the task demands in Experiment 2 (in which all sentences were presented in noise) lead to overall higher cognitive load compared to Experiment 1 cannot be determined here, but could in principle explain the lower microsaccade rate even before sentence onset. Further, previous work has shown greater cognitive load for informational compared to energetic masking (Ohlenforst et al., 2017b; Zekveld, Rudner, Kramer, Lyzenga, & Ronneberg, 2014). Moreover, behavioural performance was lower for Experiment 2, relative to Experiment 1 across the 4 conditions (Experiment 1: LAC: 91%, HAC: 93%, LAN: 90%, HAN: 86%; Experiment 2: LA6: (89%, HA6: 90%, LA0: 85%, HA0: 80%), again consistent with this task being more difficult/more demanding. These results are compatible with the idea that pupil dilation is sensitive to transient increases in task demands, whereas microsaccades are sensitive to overall demands of experimental procedures, but this would need to be explicitly tested in the future.
4.4 Pupil dilation and listening effort

Several previous studies demonstrate that the mean pupil size increases with the degree of acoustic degradation of speech and background noise (Kuchinsky et al., 2013; Ohlenforst et al., 2018, 2017b; Wendt et al., 2018; Winn et al., 2015; Zekveld et al., 2014a, 2013a, 2013b, 2010). Our data are consistent with these previous studies that have demonstrated increased pupil size related to increased acoustic demands (Zekveld and Kramer, 2014). Further, previous studies show increased cognitive load during informational (Experiment 2) compared to energetic (Experiment 1) masking (Ohlenforst et al., 2017b; Zekveld, Rudner, Kramer, Lyzenga, & Ronneberg, 2014). These data have been understood to mean that listening effort, when sentences are degraded in some way, is reflected in pupil size.

A well-established literature demonstrates that “mental effort” of many different kinds is also associated with pupil dilation (Alnaes et al., 2014; Beatty, 1982; Hornsby, 2013; Kahneman, 1973; Kahneman and Beatty, 1966; Koelewijn et al., 2012; Mathôt, 2018; McCloy et al., 2017; Papesh et al., 2012; van der Meer et al., 2010). Larger pupil size is observed for more semantically difficult words (Chapman and Hallowell, 2015), unknown vs known words (Ledoux et al., 2016), and syntactic complexity (Carroll and Ruigendijk, 2013; Gibson, 1998; Wendt et al., 2016). Moreover, the pupil reflects cognitive control which is recruited for both acoustics and linguistic demands.

Wendt et al., (2016) demonstrated pupil enlargement for complex object-verb subject sentences compared to the less complex subject-verb-object sentences. Moreover, and similar to the current study, their study tested the combined effects of syntactic complexity (complex vs simple) and background noise level directly and demonstrated that syntactic complexity and noise both influence pupillary responses, albeit in different epochs. They tested using the same type of noise as we used in Experiment 2, but at SNRs of +12 and -6 dB, higher and lower than the SNRs we used (0 and +6 dB SNR). Wendt et al (2016) also had
the noise start earlier than the sentence, perhaps engaging segregation processes that would not be engaged in our Experiment 1. Thus, by using a different noise, and starting it earlier than the target, we think we are engaging a different set of cognitive processes during listening. They observed effects of both perceptual (acoustic) and linguistic challenges, as we did. However, their results were somewhat different, in that they observed that the effect of noise was only evident for simple sentences – the added effect of noise was negligible for complex sentences. In contrast, despite using a narrower range of noise levels, we observed a clear effect of noise at both levels of ambiguity. The two challenges (semantic ambiguity and noise) appeared to of acted additively, suggesting that they are somewhat independent. The differences between our finding and that by Wendt et al. (Wendt et al., 2016) may have stemmed from differences in task demands in the two studies: the stimulus used by Wendt et al. was comprised of auditory and visual components, whereas our task was solely auditory. Another possibility is that the syntactic complexity manipulation they used is more demanding than the ambiguity manipulation that I used.

4.3 Ceiling effect or U-Shaped response?

An interesting contribution of the current study is that we show the limitations of the pupillary response when used a monotonic index for effort. Our pooled data point to a potential ‘ceiling effect’ or an ‘inverted U-shaped’ relationship between the level of task challenge and the corresponding pupil response. That is, the pupil will dilate with increasing task challenge level, and then will remain at this relative level of dilation, with little ‘room’ for further dilation or even constrict as the task becomes exceptionally more difficult. The inverse-U-shaped function of the pupil has been recently confirmed (Ohlenforst et al., 2017b). Of clinical relevance, in clinical audiology, our data suggest that perhaps a physiological plateau or ceiling effect may occur at a certain level of task difficulty as patients might disengage. The implications of this ceiling effect on pupil dilation may affect current clinical means of interpreting pupillary responses for a given task. This is highlighted by the observed decoupling between pupillary and behavioral data. Specifically,
adding noise (an acoustic dimension) to an already cognitively demanding sentence (linguistic dimension) produced additional difficulty resulting in a decrease in performance (i.e., number of correct responses in semantic relatedness task). That is, an increased cognitive load resulting from ambiguity leaves less cognitive load for processing acoustics.

Though pupillometry was sensitive enough to pick up differences between sentence type in the clear/low noise condition (behaviour data did not), this was not the case in the high noise condition. We believe the lack of differences observed in the high noise condition could potentially be described by the aforementioned ceiling effect/ inverted U-shape of the pupillary response where there is a limit to pupillary dilation, and pupillary responses are not discernable in very difficult conditions. To this end, with the additive difficulty (based on our behavioural data herein), the pupil plateaus or constricts as speech becomes very difficult. Thus, it is difficult to discern difficulty as a task becomes exceptionally difficult. Although we did not observe this ceiling clearly, it is possible that this is the reason why Wendt et al (2018) did not observe an effect of noise when sentences were syntactically complex. This presents another important consideration and limitation of pupillometry use for listening effort.

Our eye-tracker was not calibrated, and we could not provide pupil size values in mm to further elucidate the ceiling effect. Another potential limitation of our study is that we did not test a broad range of noise levels in either experiment. Perhaps, the effects of these two challenges are different at different noise levels.

Despite these experimental limitations, our results indicate that when using pupil dilation as an index of listening effort, the cognitive demands of the speech stimuli are important to consider. Semantic ambiguity is ubiquitous in language, with up to 80% of words having more than one meaning. Nevertheless, utterances containing such words pose a linguistic challenge that is reflected in pupil dilation. Although it is tempting to infer that differences in pupil dilation reflect differences
only in hearing ability, our data do not support this. The nature of the stimuli matter, as would individual differences in the cognitive abilities recruited to compensate for both linguistic and perceptual demands may manifest as differences in pupil dilation. Only by recognizing that different challenges to comprehension are met through the recruitment of different cognitive abilities that may differ among individuals, will we be able to understand individual variability in effortful listening, which will enable individually tailored interventions to minimize effort.

5 Conclusion
Further to the recognized pupil dilation observed for both types of background noise, we extend previous research by showing an effect of semantic ambiguity. These two effects seem to be additive, at least at the noise levels that we used. Overall, our findings suggest that when indexing listening effort, the ubiquity of ambiguous words pose a confounding factor that is reflected in pupil dilation. Thus, individual differences in cognitive abilities recruited to cope with the acoustic and linguistic challenges will manifest as differences in the pupillary response. These findings invite future assessment methods of listening effort to control for ambiguous words and microsaccadic responses in an effort to reduce heterogeneity in hearing outcomes and improve clinical diagnoses.
References


Dalmaso, M., Castelli, L., Scatturin, P., Galfano, G., 2017. Working memory load modulates microsaccadic rate. J. Vis. 17, 6. https://doi.org/10.1167/17.3.6


Appendices

Appendix A: Ethics approval

Date: 15 May 2019
To: Ingrid Johansen
Project ID: 106470
Study Title: Electromyography and behavioral studies of speech and sound perception
Application Type: Continuing Ethics Review (CER) Form
Review Type: Delegated
REB Meeting Date: 21/May/2019
Date Approval Issued: 15/May/2019
REB Approval Expiry Date: 02/Jul/2020

Dear Ingrid Johansen,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonization Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 5 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00005948.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Daniel Wierzycki, Research Ethics Coordinator, on behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).
Appendix B: Letter of information and consent

Behavioral studies of speech and sound perception

Principal Investigator:
Dr. Ingrid S. Johnsrude
Department of Psychology, The University of Western Ontario, London, ON
Telephone: [redacted]

Introduction
You are being invited to participate in a research study about human perception of speech and sound because you are older than 18 (or 17 if first year student) and have no hearing difficulty or neurological problems. The purpose of this study is to investigate how humans perceive sound and speech, and how sounds and speech might change our experience or memory of sounds or sights.

The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research. It is important for you to understand why the study is being conducted and what it will involve. Please take the time to read this carefully, and feel free to ask questions if anything is unclear or if there are words or phrases you do not understand.

Research Procedures
The experiments conducted as part of this study will test how humans hear, see, remember, and behave when they listen to sounds (including speech or music) or see visual stimuli. If you agree to participate, you will be asked to listen to auditory stimuli. You might also be asked to perform a task testing, for example, discrimination/detection, memory, or attention while you are listening to the auditory stimuli. Tasks could be related or unrelated to the auditory stimuli and may include watching a movie, making simple responses about whether you detect the presence of or differences between stimuli, to tap in time with the stimuli, give verbal or written feedback, and/or to make ratings about your impressions of the stimuli. General cognitive and hearing abilities may additionally be assessed. Eye movements may be recorded as an additional behavioral measure in response to experimental stimuli (using a camera capturing directional eye movements). It is anticipated that the entire task will take no more than 4.5 hours. The task(s) will be conducted in the Brain and Mind Institute in the Western Interdisciplinary Research Building, the Social Sciences Building, the National Centre for Audiology, or the Roberts Research Institute on the University of Western Ontario campus. There will be a total of 1,000 participants.

Inclusion and Exclusion Criteria
Individuals who are at least 17 years of age having hearing and vision adequate to perform the task are eligible to participate in this study. Where critical for the experiment, individuals may also undergo a brief audiometric assessment of hearing abilities. Individuals who are younger than 17 years of age, have any neurological diseases in their past, or who have hearing damage or vision problems too severe to complete the task will be excluded from the study. Where speech stimuli are used, individuals being left handed may be excluded from the study.

Risks and Benefits
There are no known or anticipated risks or discomforts associated with participating in this study. Some individuals may temporarily experience light fatigue when eye movements are recorded because you will be asked to reduce body movements. Although you may not directly benefit from participating in this study, the information gathered may provide benefits to society as a whole which include enhancing our scientific
understanding of auditory perception and hearing, leading to advancements in medical care (for example, hearing aids) to counteract decline in hearing often accompanying healthy aging.

It is possible that, in the case of a brief hearing assessment, a previously unknown hearing impairment is identified. If this were to occur, we will encourage you to seek professional assessment from your family practitioner or audiologist. Please note that the hearing assessments that may be included in this experiment are not clinical tests and cannot be used to make any diagnosis. We are, however, able to provide information about obtaining an assessment at the UWO audiology clinic in Elborn College.

Compensation
If you were recruited via SONA, you will receive the following compensation: If you are a Psychology 1000 student you will receive 0.5 research credits per half hour of your participation (as pro-rated compensation). If you are an other-than-Psychology 1000 student you will receive compensation based on information provided in your course outline. If you were recruited via other means than SONA you will receive monetary compensation for your participation in this study ($5 per half hour of participation). If you do not complete the entire study you will still be compensated a pro-rated amount.

Voluntary Participation
Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time (which includes withdrawing your data) with no effect on your future academic status.

Confidentiality
Any information obtained from this study will be kept confidential. In the event of publication, any data resulting from your participation will be identified only by case number, without any reference to your name or personal information. The data will be stored on a secure computer in a locked room. Both the computer and the room will be accessible only to the investigators. After completion of the experiment, data will be archived on storage disks and stored in a locked room. Any documents identifying you by name will be kept separately from your data, and will be destroyed after 7 years.

Following publication of results, the study data may be made openly available to other researchers on public servers (e.g., open science framework, BIDS, or Dryad Digital Repository). This is done in an effort to enhance transparency of research results, advance knowledge, avoid research duplication and maximize research benefits to the general population. None of your personal information will be shared, only anonymized data will be made publicly available.

Representatives of the University of Western Ontario Health Sciences Research Ethics Board may require access to your study-related records or may follow up with you to monitor the conduct of the study.

Contacts for Further Information
If you would like to receive a copy of the overall results of the study, or if you have any questions about the study please feel free to contact the Principal investigator at the contact information provided above.

If you have any questions about your rights as a research participant or the conduct of the study you may contact:

The Office of Research Ethics
The University of Western Ontario

This letter is yours to keep for future reference.
**Consent Form**

**Project Title:** Behavioral studies of speech and sound perception  
**Study Investigator’s Name:** Dr. Ingrid S. Johnsrude

I have read the Letter of Information, have had the nature of the study explained to me and I agree to participate. All questions have been answered to my satisfaction.

Participant’s Name (please print): ________________________________

Participant’s Signature: ________________________________

Date: ________________________________

Circle as appropriate:
I am **interested**/am **not interested** in being contacted so that I may participate in other studies conducted by the researcher and their lab group.

_________________________________________________________

*Please leave this part blank for the experimenter to complete*

Person Obtaining Informed Consent (please print): ________________________________

Signature: ________________________________

Date: ________________________________
Curriculum Vitae

<table>
<thead>
<tr>
<th>Name:</th>
<th>Mason Kadem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-secondary Education and Degrees:</td>
<td>Western University London, Ontario, Canada B.Sc. H</td>
</tr>
<tr>
<td></td>
<td>Western University London, Ontario, Canada MSc</td>
</tr>
</tbody>
</table>