8-19-2022 1:00 PM

Listening effort: Separating the subjective from the objective

Joseph E. Rovetti, *The University of Western Ontario*

Supervisor: Johnsrude, Ingrid S., *The University of Western Ontario*

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

Abstract

Challenges such as background noise may increase "listening effort." This construct has been operationalized as the recruitment of cognitive resources during listening (objective effort) or as the self-reported feeling of effort (subjective effort). In the current study, I compared these two dimensions of listening effort directly. Normal-hearing adults listened to highly intelligible passages across several signal-to-noise ratios (SNRs), with reaction time on a secondary task (objective effort) and effort ratings (subjective effort) measured in separate blocks. As the SNR became less favourable, subjective effort appeared to increase continuously, while objective effort only began to increase at a much less favourable SNR. This suggests that although listening effort increases with cognitive demand, these two dimensions may respond differently. However, the greater responsiveness of subjective effort may be due to participants rating the difficulty rather than effort. Further, listening effort appeared to increase before speech intelligibility decreased, suggesting that effort helps maintain intelligibility.

Keywords

Listening effort, Speech perception, Speech-in-noise, Dual-task paradigm, Self-report, Hearing loss.

Summary for Lay Audience

Have you ever struggled to have a conversation in a noisy environment like a busy café? If so, you likely experienced "listening effort"—the phenomenon of working hard to understand speech. Quite often, people complain of high listening effort even if they can understand speech well, and even if they appear to have normal hearing. Ideally, clinicians would be able to measure listening effort when assessing patients, but no standard method currently exists. This is in part due to the confusion around the definition of listening effort. Some researchers consider it how hard the brain is working to listen (which I call objective listening effort), while others consider it how hard listeners feel like they are working (which I call subjective listening effort). In fact, many argue that these are two distinct dimensions of listening effort. In this study, I compare objective and subjective effort directly. To do this, I had participants with normal hearing listen to many speech passages with noise in the background, which varied from very favourable to unfavourable. In one half of the experiment, I used a cognitive task to measure participants' objective effort, and in the other half, I used a questionnaire to measure subjective effort. I found that as the intensity of the noise increased, subjective effort (participants' feeling of effort) increased continuously. In contrast, objective effort (how hard participants' brains were working) started fairly stable and only began to increase at a much higher noise level. This suggests that both dimensions of listening effort increase as listening difficulty increases, but that they may respond differently. In particular, it appears that subjective effort can be high even when objective effort is not. This was a surprising result, and it may suggest that participants were actually rating how hard the task was rather than the effort they felt. In addition, listening effort appeared to increase before speech understanding dropped, which suggests that people may invest listening effort as a way to keep speech understandable. More knowledge of listening effort may help clinicians to diagnose and treat more cases of hearing loss.

Co-Authorship Statement

This research will likely be submitted for publication, with Joseph Rovetti as the first author. The co-authors will be as follows: Jaimy Hannah, Stephen Van Hedger, Hasan Saleh, Susan Scollie, and Ingrid Johnsrude. Jaimy Hannah contributed to the design and programming of the experiment as well as the statistical analyses. Stephen Van Hedger contributed to the modification of the stimuli and the statistical analyses. Hasan Saleh and Susan Scollie provided the unmodified version of the stimuli. Ingrid Johnsrude served as the principal investigator, provided funding, contributed to the conceptualization and design of the experiment, to the interpretation of results, and to the editing of this thesis.

Acknowledgments

My deepest of gratitude goes to my supervisor, Ingrid Johnsrude. As far as mentors go, I couldn't have asked for a better scientist or a better human being. Thanks for always finding a way to keep my spirits high. This appreciation extends to my supervisory committee, Blake Butler and Ken McRae, as well as the other scientists who contributed to this project. Thank you for your constructiveness and kindness, both of which I needed.

I also thank the rest of the CoNCH Lab and other members of the Brain and Mind Institute for their advice and feedback, especially Vanessa Irsik, Bruno Mesquita, Olivia Richards, and Rebekka Lagacé-Cusiac. Most of all, thank you Jaimy Hannah. Without your help, this project would be much less good, and I would be much less well.

Finally, I couldn't have completed this thesis without the support of my friends and family to get me through a challenging two years. To my dad, you inspire me every day. I wouldn't be where I am now without the sacrifices you've made. Thank you.

Table of Contents

Abstract ii

Summary for Lay Audience iii

Co-Authorship Statement iv

Acknowledgments v

Table of Contents vi

List of Tables viii

List of Figures ix

List of Appendices x

Chapter 1 1

1 Introduction 1

1.1 The importance of listening effort 1

1.2 The varying definitions of listening effort 3

1.3 Objective listening effort 5

1.4 Subjective listening effort 7

1.5 Comparing objective and subjective effort 9

1.6 The current study 12

Chapter 2 15

2 Methods 15

2.1 Participants 15

2.2 Materials 15

2.2.1 Speech stimuli 15

2.2.2 Secondary task 17

2.2.3 Effort rating scale 17

2.2.4 Comprehension questions 17

2.3 Design 18

2.4 Procedure 19

2.5 Data quality 23

2.6 Data analysis 24

2.6.1 Pre-processing 24

2.6.2 Statistical analysis 26

Chapter 3 28

3 Results 28

3.1 Gist understanding 28

3.2 Secondary-task reaction time 30

3.3 Effort rating 31

3.4 Comparing measures 33

Chapter 4 36

4 Discussion 36

4.1 Objective effort increased as the SNR became less favourable 36

4.2 Subjective effort responded before objective effort 41

4.3 Strengths, limitations, and future directions 45

4.4 Conclusion and implications 47

References 49

Appendices 68

Curriculum Vitae 85

List of Tables

Table 1: Connected Speech Test passage lists 18

List of Figures

Figure 1: Gist understanding by SNR.. 29

Figure 2: Secondary-task reaction time by SNR.. 31

Figure 3: Effort rating by SNR 32

Figure 4: Functions for all measures.. 34

List of Appendices

Appendix A: Connected Speech Test passage sentences 68

Appendix B: Passage comprehension questions 78

Appendix C: Ethics approval 79

Appendix D: Materials for the secondary task 80

Appendix E: Materials for the effort rating scale 81

Appendix F: Secondary-task reaction time by letter position 82

Appendix G: Analysis of comprehension 83

Appendix H: Histogram of RAPM scores 84

Chapter 1

1 Introduction

Speech is one of the most remarkable of human abilities, as well as one of the most unique. At a physical level, speech is little more than air being sent from the lungs through the vocal folds to produce pressure waves. And yet, it is the ability of articulators such as the lips and tongue to manipulate these waves, and the ability of the brain to decode them into meaning, that enables the majority of human communication. Even more remarkable is the incredible resilience of this ability. For instance, speech can be understood even if background noise is present (e.g., in a busy café) or if the listener has hearing loss. However, these challenges come at two costs for the listener. First, and most obviously, speech will not always be understood correctly. Second, even to the extent that speech can be understood, doing so under adverse conditions is associated with greater effort. This latter cost has been referred to as “listening effort,” defined most broadly as the phenomenon of working hard to listen (e.g., to speech). Despite an abundance of research, there is considerable debate as to how listening effort should be defined and measured. In the current study, I experimentally compared what are proposed to be two distinct dimensions of listening effort: the process of recruiting cognitive resources during listening (objective listening effort) and the feeling of listening being effortful (subjective listening effort).

1.1 The importance of listening effort

The first use of the term “listening effort” with respect to speech and hearing was in a society proceedings from the 1920s (Berry et al., 1925). This publication spoke of how deaf individuals experience elevated listening effort, and how they may disengage from listening if this effort is not sufficiently rewarding for them. However, the term only began to be used widely starting in the 1980s (Downs, 1982). Its popularity rose further in the 21st century with the emergence of cognitive hearing science, a field that emphasizes the role of cognition in hearing and speech perception (Arlinger et al., 2009).

At the time of writing, a Google Scholar search for listening effort returned just under 7000 results. The number of results published in 2021 (920) was more than double the number from five years earlier (451).

The popularity of listening effort has been driven by its substantial clinical promise. In particular, listening effort can help to account for behavioural differences that conventional hearing assessments cannot (McGarrigle et al., 2014; Herrmann & Johnsrude, 2020a). Hearing assessments usually consist of pure-tone audiometry, in which patients are tested on their ability to hear quiet pure tones (i.e., simple sounds with a sinusoidal waveform) across several frequencies and in each ear (Walker, 2013). According to this test, approximately 40% of Canadians over the age of 60 have hearing loss (Feder et al., 2015). However, pure-tone audiometry does not necessarily predict speech perception ability (Tremblay et al., 2015). Indeed, patients with normal pure-tone detection often report everyday communication difficulties (Parthasarathy et al., 2020), including elevated listening effort (Gatehouse & Noble, 2004; Hornsby et al., 2016; Pichora-Fuller et al., 2016). Indeed, even among those with diagnosed hearing loss, hearing aids are often not worn consistently because of their inability to reliably reduce listening effort (Ohlenforst et al., 2017a). Thus, to gain a complete picture of patients' hearing health, it is not enough to use performance-based measures; rather, the effort required to achieve this performance must also be considered.

Whether hearing loss decreases pure-tone detection, or only increases listening effort, depends on the cause. A wide array of pathologies can contribute to hearing loss, including ineffective transmission of sound through the outer and middle ear, degradation of hair cells in the cochlea of the inner ear, or impaired sound processing to and within the brain (Cunningham & Tucci, 2017). However, only a subset of these pathologies—generally those based earlier in the auditory pathway—lead to reduced pure-tone detection. Other pathologies, rather than interfering with the *intensity* of sound, instead interfere with the *quality* of sound (Plack et al., 2014). This reduced sound quality does not usually interfere with pure-tone detection, but it may increase the effort experienced during listening, particularly in noisy or otherwise-challenging conditions (Pienkowski, 2017). However, there is currently no accepted measure of listening effort for clinical use

(McGarrigle et al., 2014; Pichora-Fuller et al., 2016). As a result, these cases of "hidden" hearing loss remain undiagnosed and untreated. Untreated hearing loss place a substantial burden on the health care system and decreases worker productivity, contributing to its global cost of over $1 trillion per year (McDaid et al., 2021).

Elevated listening effort, whether due to challenging listening conditions or hearing loss, is associated with negative cognitive and affective consequences for the listener. On the cognitive side, Kahneman's (1973) Capacity Model of Attention states that cognitive resources are finite and can be depleted by demanding tasks, including speech perception under challenging conditions (see Pichora-Fuller et al., 2016). Thus, challenging speech perception can impair performance on other, simultaneous tasks. This includes general multitasking (Anderson Gosselin & Gagné, 2011) as well as the ability to rehearse and encode what is being heard, resulting in memory deficits (Rabbitt, 1991; McCoy et al., 2005). For instance, Rudner et al. (2018) found that, in a classroom setting, children's listening comprehension was reduced by even low levels of background noise. On the affective side, effortful listening is also associated with mental fatigue and distress (Hétu et al., 1988; Hornsby, 2013). This can motivate listeners to withdraw from social situations, which may lead to negative health outcomes or accelerate cognitive decline (Nicholson, 2009; Lin et al., 2013; Pichora-Fuller et al., 2015).

1.2 The varying definitions of listening effort

Despite the popularity of listening effort as a research topic, and its clinical promise, it continues to elude definition and measurement. As recently pointed out by Strand et al. (2020), this confusion is apparent when reviewing the literature on listening effort. Among the most-cited papers are titles such as "Listening effort and fatigue: What exactly are we measuring?" (McGarrigle et al., 2014) as well as "Listening effort: Are we measuring cognition, affect, or both?" (Francis & Love, 2020). In addition, listening effort is often conflated with terms such as "mental effort" (Panico & Healey, 2009), "perceptual effort" (Tun et al., 2009), "cognitive effort" (Obleser et al., 2012), and "cognitive load" (Zekveld et al., 2011).

As mentioned, listening effort can be broadly defined as the phenomenon of working hard to listen. The choice of the word “phenomenon” was intentional, as researchers cannot even agree whether listening effort refers to a *process* (i.e., the recruitment of cognitive resources) or a *feeling* (i.e., the subjective state of difficulty). Some reviews define listening effort as a process (McGarrigle et al., 2014; Pichora-Fuller et al., 2016; Peelle, 2018). For instance, McGarrigle et al. (2014) described it as “the mental exertion required to attend to, and understand, an auditory message” (pg. 2), and Pichora-Fuller et al. (2016) as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task” (pg. 10S). In contrast, others define it as a feeling (Johnsrude & Rodd, 2016; Herrmann & Johnsrude, 2020a). For example, Herrmann and Johnsrude (2020a) restricted their view of listening effort to “a person’s experience during listening” (pg. 2). Many studies also measure it using rating scales, which implies such a definition (Alhanbali et al., 2017; Krueger et al., 2017).

In studies of listening effort, it is usually conceptualized as one of the two definitions stated above—either as the process of effort (McGarrigle et al., 2014; Pichora-Fuller et al., 2016; Peelle, 2018) or as the feeling of effort (Johnsrude & Rodd, 2016; Herrmann & Johnsrude, 2020a). However, it has also been used as an umbrella term to refer to both definitions (Lemke & Besser, 2016; Francis & Love, 2020). Even among studies to make a distinction between the process and the feeling of effort, the terminology is not consistent. The process of listening effort has been referred to as “processing effort” (Lemke & Besser, 2016), “exerted effort” (Francis & Love, 2020), and “objective listening effort” (Picou et al., 2017), while the feeling of listening effort has been referred to as “perceived effort” (Lemke & Besser, 2016), “assessed effort” (Francis & Love, 2020), and “subjective listening effort” (Picou et al., 2017). In the current study, I borrow the terminology of Picou et al. (2017), referring to the recruitment of cognitive resources to support listening (the process) as “objective listening effort” and the subjective state of difficulty during listening (the feeling) as “subjective listening effort.

1.3 Objective listening effort

Objective effort is the recruitment of cognitive resources to support listening. It likely arises, primarily, from the interaction between cognitive demand and motivation (Pichora-Fuller et al., 2016; Strauss & Francis, 2017; Peelle, 2018). Listeners will recruit more cognitive resources as the cognitive demand of a listening situation (e.g., the level of background noise) increases, assuming they are motivated to do so (Eckert et al., 2016; Matthen, 2016). Motivation may in practice be a binary yes-no switch (Herrmann & Johnsrude, 2020a), or it may have varying degrees (Pichora-Fuller et al., 2016; Peelle, 2018). In addition to demand and motivation, objective effort likely also depends on factors such as the cognitive and perceptual profile of listeners (Lemke & Besser, 2016). For instance, listeners with lower cognitive processing efficiency or with hearing loss (e.g., older adults) may require more cognitive investment to meet demands, even when motivation is constant (Reuter-Lorenz et al., 1999; Pichora-Fuller, 2003; Mattay et al., 2006). Eventually, though, as the effective signal-to-noise ratio (SNR) becomes less favourable, the cognitive demand will become too great for all listeners (Herrmann & Johnsrude, 2020a). If listeners remain highly motivated, objective effort may plateau once their capacity limit is reached; and if listeners lose motivation, they may disengage and their objective effort will drop off sharply at or even before this limit (Zekveld & Kramer, 2014; Ohlenforst et al., 2017b; Ayasse & Wingfield, 2018; Wendt et al., 2018).

The specific cognitive resources recruited probably depend on the listening challenges present (Johnsrude & Rodd, 2016; Lemke & Besser, 2016; Peelle, 2018; Herrmann & Johnsrude, 2020a). Competing speech or other forms of background noise, for instance, interfere with speech perception in two primary ways. First, energetic masking occurs when noise occludes the target speech, rendering it less available due to physical interactions in the periphery (Fletcher & Galt, 1950; Pollack, 1975). Second, informational masking—broadly defined as any masking that is not energetic—usually occurs when target and noise are both audible but difficult to separate perceptually (Watson et al., 1976; Brungart, 2001). These two forms of masking place a load on cognitive resources that support perceptual closure to “fill in” any portions of the target speech that were occluded by noise (e.g., based on linguistic or general world knowledge), stream segregation to separate the target speech from noise (e.g., based on spectral cues), and selective attention to the target speech over the noise (Mattys et al., 2012; Johnsrude & Rodd, 2016; Rovetti et al., 2022a).

Objective effort is most commonly assessed using behavioural measures. These measures include single-task paradigms, in which a listener’s reaction time to a listening task (e.g., identifying the final digit in a series of three digits) is used to measure listening effort (Houben et al., 2013), with slower reaction times thought to indicate greater effort. The most widespread behavioural measure of objective effort is the dual-task paradigm (Rabbitt, 1966; Gagné et al., 2017). In this paradigm, participants complete a “primary” listening task while also completing a unrelated “secondary” task. The secondary task is generally concurrent, such as responding to a simple visual probe (Hornsby, 2013) or classifying stimuli (Rodd et al., 2010; Brown et al., 2020) during listening. The secondary task can also be sequential, such as later recalling digits presented during the primary task (Picou et al., 2011). The dual-task paradigm assumes that listeners’ cognitive capacity is finite, and that the more cognitive resources are used for a primary listening task, the fewer will be left over for the effective completion of the secondary task (Kahneman, 1973; McGarrigle et al., 2014). Thus, poorer performance on the secondary task is thought to indicate greater objective effort during the primary task. For concurrent secondary tasks, reaction time is generally the favoured measure of performance given its greater sensitivity (Baer et al., 1993), although task accuracy can yield similar results (Anderson-Gosselin & Gagné, 2011). For this paradigm to be effective, listeners should prioritize the primary task over the secondary task and dedicate all of their cognitive resources to these two tasks at all times (Gagné et al., 2017).

Objective effort is also assessed using physiological measures. Increased cognitive resource recruitment elevates stress levels, which is reflected in increased sympathetic activity and decreased parasympathetic activity (Staal, 2004). These effects drive several peripheral physiological changes that have been used to measure the stress from objective effort, including increased skin conductance and reduced heart rate variability (Mackersie & Cones, 2011; Mackersie et al., 2015). The most common peripheral physiological measure is pupil size (Beatty & Kahneman, 1966; Winn et al., 2018). A larger pupil size is often interpreted as greater objective effort, since this response is regulated by the same brain area (the locus coeruleus) that regulates attention and cognitive control (Raizada & Poldrack, 2008; Joshi et al., 2016). Less commonly, objective effort has also been operationalized as increased activation of relevant brain areas. The increased brain activity associated with listening in noise can be measured directly using electrophysiological methods such as electroencephalography, with indices including alpha power oscillations (Obleser et al., 2012), amplitude of the N1 ERP (Obleser & Kotz, 2011), and wavelet phase synchronization stability of the late auditory response (Bernarding et al., 2013). Brain activity also causes blood flow to the active areas via neurovascular coupling (Glover, 2011). Thus, objective effort has also been assessed using hemodynamic measures such as functional magnetic resonance imaging (Wild et al., 2012; Ritz et al 2022) and functional near-infrared spectroscopy (Wijayasiri et al., 2017; Rovetti et al., 2022a).

1.4 Subjective listening effort

Subjective effort is the feeling of effort during listening. The relationship between objective and subjective effort is unclear. For instance, subjective effort may arise as a consequence of objective effort (Johnsrude & Rodd, 2016), or it may have the potential to be elevated even when objective effort is not (Lemke & Besser, 2016). According to the Model of Listening Engagement, “[subjective effort is] the consequence of the recruitment of cognitive and other resources [i.e., objective effort]” (Herrmann & Johnsrude, 2020a, pg. 3). In particular, subjective effort may only be felt when listeners’ available cognitive resources are inadequate or barely adequate to meet the demands of a situation (Johnsrude & Rodd, 2016; Herrmann & Johnsrude, 2020a; but see Strauss & Francis, 2017 for an alternative model). In this case, subjective effort would be—as with objective effort—dependent on the cognitive and perceptual profile of listeners. Cognitive capacity is also not a unitary construct; rather, listeners have an assortment of cognitive abilities that may contribute to speech perception (e.g., working memory, attention), each with a different limit (Wickens, 2008; Van Hedger & Johnsrude, 2022). If any one of these abilities is overly taxed, even if the others are not, subjective effort may be felt. Capacity limits also vary over time given the personal state of the listener, such as whether they are tired (Wright, 2014; Richter et al., 2016). In addition, rather than following instantaneously from objective effort, subjective effort may only manifest after prolonged resource recruitment (Lemke & Besser, 2016; Bain et al., 2020).

Subjective effort is usually assessed using self-report measures. Self-report measures involve asking the listener to report the effort that they feel during listening (Feuerstein, 1992). This approach assumes that listeners can accurately perceive, remember, and report the feeling of effort (McGarrigle et al., 2014). Listeners are generally asked to rate effort after listening to one stimulus, such as a sentence or passage (Picou et al., 2011; Holmes et al., 2018), or after multiple stimuli (Fraser et al., 2010; Mackersie & Cones, 2011; Rudner et al., 2012). However, this reporting can also occur in real time, with listeners able to update their rating when their feeling of effort changes. Effort ratings are usually collected using one-item visual analog scales, a continuous one-dimensional scale with two endpoints (Hayes & Patterson, 1927). These scales may consist of an uninterrupted line (McAuliffe et al., 2012) or have marks to indicate different levels of effort (Brons et al., 2014; Rennies et al., 2014; Rudner at el., 2012; see Lau et al., 2019). Two of the most common visual analog scales are the NASA Task Load Index (in particular the “Effort” scale; Hart & Staveland, 1988) and the Visual Analog Scale of Fatigue (Lee et al., 1991), both adapted from their original uses. More complex, multidimensional questionnaires have also been widely used, such as the 49-item Speech, Spatial and Qualities of Hearing Scale (Gatehouse & Noble, 2004).

To avoid possible response bias (see Moore & Picou, 2018), subjective effort has also been assessed via more creative means. Some visual analog or other scales have asked not about “effort” or a similar term, but rather a preference or behaviour thought to be more accessible than the feeling of effort. For instance, in adults, Picou et al. (2017) and Picou and Ricketts (2018) asked participants to rate how much they wanted to improve the listening situation, how much they wanted to give up, and how tired they felt. In children, Picou et al. (2019) also included questions about how much participants wanted to increase the speaker’s volume or how long the task felt, with the latter assuming that time is perceived as passing slower when cognitive demand is higher (Block et al., 2010; Sucala et al., 2010). In addition, McLaughlin et al. (2021) attempted to use a discounting task to approximate participants' subjective effort behaviourally. In this task, participants were given a choice between completing a more difficult listening task for a larger financial reward or an easier task for a smaller reward (see Westbrook et al., 2013). In this paradigm, the larger the reward that a participant was willing to forgo to avoid the more difficult listening task, the higher their subjective effort was assumed to be.

1.5 Comparing objective and subjective effort

Listening effort is an abstract construct with no physical reality. Despite this, measures of listening effort are often conflated with listening effort itself. In this sense, listening effort has been subjected to the reification fallacy, in which an abstract idea is treated as if it were a concrete object (Whitehead, 1925). This view of listening effort opens the door for other forms of fallacious reasoning. For instance, the various measures of listening effort are often assumed to be interchangeable—different approaches to assessing the same construct (but see Herrmann & Johnsrude, 2020a for a critique). However, in practice, these measures frequently disagree with one another. This conflation is a case of the jingle fallacy, in which it is falsely assumed that two things sharing the same name are necessarily equivalent (Thorndike, 1904; see Strand et al., 2020). Indeed, across studies, different measures have yielded inconsistent conclusions as to the factors that affect listening effort. For instance, previous research has disagreed on whether visual speech cues (e.g., seeing the face of the speaker) affects listening effort (see Brown & Strand, 2019). This appears to depend on the measure used, with secondary-task reaction time often revealing that visual cues increase listening effort (Brown & Strand, 2019; Fraser et al., 2010; Anderson Gosselin & Gagné, 2011) and dual-task recall accuracy revealing a decrease in listening effort (Rudner et al., 2016; Sommers & Phelps, 2016). Inconsistencies have also been reported between objective and subjective measures of listening effort. All studies measuring subjective effort have reported that it decreases with the addition of context (Johnson et al., 2015; Holmes et al., 2018). In contrast, all studies that have reported similar levels of effort or increased effort with the addition of context have measured objective effort (Tun et al., 2009; Desjardins & Doherty, 2014; Lau et al., 2019; Borghini & Hazan, 2020).

When multiple measures of listening effort are used in the same study, they may or may not produce consistent results. In studies assessing objective and subjective effort, the same pattern of results is often reported (Picou et al., 2011; Koelewijn et al., 2012; Wu et al., 2016; Wendt et al., 2016; Dimitrijevic et al., 2019). For instance, in Koelewijn et al. (2012), young and middle-aged adults listened to speech presented in three types of background noise (stationary, fluctuating, and single-talker) presented at two levels (84% and 50% intelligibility). Participants' reported the most subjective effort, and exhibited the largest pupil response, with a single-talker masker and in the 50% intelligibility conditions. However, other studies have reported different results for objective and subjective effort (Feuerstein, 1992; Hicks & Tharpe, 2002; Larsby et al., 2005; Anderson Gosselin & Gagné, 2011; Hornsby, 2013; Pals et al., 2013; Mackersie et al., 2015; Picou & Ricketts, 2018; Lau et al., 2019; Carolan et al., 2021). For instance, consistent with the Model of Listening Engagement (Herrmann & Johnsrude, 2020a), the majority of studies have found that objective effort may be more responsive to cognitive demand (Downs & Crum, 1978; Hicks & Tharpe, 2002; Hornsby, 2013; Pals et al., 2013; Mackersie et al., 2015). In Hicks and Tharpe (2002), children with and without hearing loss listened to speech in challenging conditions. Compared to a baseline, children with hearing loss had slower secondary-task reaction times than children with normal hearing, but subjective effort did not differ. Further, Hornsby et al. (2013) had middle-aged and older adults with hearing loss listen to speech with and without hearing aids. Secondary-task reaction time and recall accuracy indicated that hearing aids decreased objective effort, while there was no such effect on subjective effort.

Even when the pattern of results for objective and subjective effort is similar, these measures rarely correlate well (Zekveld et al., 2010; Zekveld et al., 2011; Johnson et al., 2015; Picou et al., 2017; Rovetti et al., 2019; McGarrigle et al., 2020), despite some notable exceptions (Koelewijn et al., 2012; Dimitrijevic et al., 2019). For instance, in Johnson et al. (2015), adults with normal hearing listened to low-predictability and high-predictability sentences at SNRs of +2 dB, 0 dB, -2 dB, and -4 dB. Self-reported and recall-based measures both suggested that listening effort was higher for low-predictability sentences and less favourable SNRs. However, when comparing these measures within each condition, their correlations were weak (*r*s = -.21–.14) and nonsignificant. In response to such results, two recent studies assessed the correlation between several measures of listening effort, including self-report, behavioural, and physiological. The first of these studies, Strand et al. (2018), tested seven measures of listening effort and determined that their average correlation was weak ($r \approx .22$). This was consistent with the later findings of Alhanbali et al. (2019), which reported a similarly weak average correlation ($r \approx .16$). The authors speculated that different measures may tap into different dimensions of listening effort rather than assessing the same, unitary construct. Other researchers have also speculated that listening effort is multidimensional, with the most common distinction made being between objective and subjective effort (Hornsby, 2013; Johnson et al., 2015; Lemke & Besser, 2016; Lau et al., 2019; Francis & Love, 2020; Herrmann & Johnsrude, 2020a).

Existing studies comparing objective and subjective effort suffer from two common limitations. First, most of this research uses speech that is too challenging to be highly intelligible across conditions (Koelewijn et al., 2012; Johnson et al., 2015). This is in contrast to the SNRs experienced in everyday life, which are typically favourable enough to allow normal-hearing listeners to achieve very high intelligibility (Smeds et al., 2015; Wendt et al., 2016). The use of very challenging SNRs increases the chances that both objective and subjective effort will be elevated from baseline levels. As a result, it is difficult to detect any differences in the response profiles of objective and subjective effort. For instance, it is unclear whether, as cognitive demand (e.g., background noise) increases, one of these two dimensions of listening effort may begin to increase before the other. Such a divergence would likely occur at relatively favourable SNRs, since listening effort typically begins to increase while speech is still fully intelligible (Houben et al., 2013). However, although studies have investigated the point at which cognitive demands become too great, few have considered when listening effort begins to increase in the first place. In addition, by using speech that is too challenging, there is a risk that participants will disengage (Wu et al., 2016; Pichora-Fuller et al., 2016) or that their performance may bias effort ratings (Moore & Picou, 2018; Kahneman & Frederick, 2002). Second, most existing studies rely on very large and often binary manipulations of cognitive demand (Hornsby, 2013; Fraser et al., 2010; Holmes et al., 2018). Even factors that can be manipulated very finely rarely are, with studies commonly using SNR steps of

5 dB or greater (Wendt et al., 2016; Lau et al., 2019). These large manipulations leave little room to detect fine-grained differences between objective and subjective effort, including their response to cognitive demand.

One notable exception to these limitations was Wu et al. (2016), in which younger adults with normal hearing listened to sentences between the SNRs of +10 dB and -10 dB in 2-dB steps. As participants listened, they completed either a simple or complex secondary task in a dual-task paradigm, with reaction time on these tasks serving as measures of objective effort. After a block of 20 sentences, they also used an adapted version of the NASA Task Load Index to rate subjective effort. The SNRs used could not reveal the point at which listening effort began to increase, but rather where it peaked. The two secondary tasks yielded similar results, with objective effort peaking around the SNRs of 0 dB and -2 dB. At more favourable SNRs, objective effort was likely lower because it was not as necessary; and at less favourable SNRs, it was likely lower because participants reached a point of cognitive overload. Subjective effort exhibited the same peak at the same SNR, around -2 dB, although no direct comparison was made between these two dimensions of listening effort. These results suggest that objective and subjective effort respond similarly to cognitive demand. However, the results of Wu et al. (2016) were also limited. Crucially, this study had participants rate subjective effort during the same blocks in which they were also completing the secondary tasks (see Anderson Gosselin & Gagné, 2011; Picou & Ricketts, 2018). When using this design, it is unclear if participants are able to separate the specific subjective effort of the primary task from the overall effort of both tasks combined (Yeh & Wickens, 1988; Recarte et al., 2008). If participants are able to do so, they most likely rely on more conscious processing, rather than the feeling of effort that these measures aim to capture. Their perceived performance on the secondary task may also have the potential bias their effort ratings. As a result, the findings of Wu et al. (2016) should be interpreted with caution.

1.6 The current study

Existing studies measuring both objective and subjective effort are insufficient to compare the response profiles of these constructs to cognitive demand. As a result, it is unclear how objective and subjective effort may respond differently as cognitive demand increases (e.g., whether one dimension of listening effort may increase before the other). To address this, I recruited a large sample of normal-hearing adults (n = 224) to complete an online study. Participants completed two blocks of testing, each presenting them with 14 speech passages. Within each block, each passage was assigned a different SNR. These SNRs were manipulated from +11 dB to -1 dB in 1-dB increments (as well as a clear condition). Pilot testing confirmed that all but the most challenging SNRs were highly intelligible. In one block, while listening to the passages, participants also completed a secondary task in which they judged whether letters presented on the screen were uppercase or lowercase. Their reaction time on this secondary case-judgement task served as a measure of objective effort. In the other block, while listening to the passages, participants used a visual analog scale to provide effort ratings continuously, with their average passage rating serving as a measure of subjective effort. After each passage, participants' also reported whether they understood the gist of the passage, which served as an index of speech intelligibility. The design of the current study allowed for the response profiles of objective and subjective effort to be compared directly. It also avoided most limitations of previous studies, including the SNRs being too challenging, manipulations being too large, as well as effort ratings being made after completing a dual-task block.

As the SNR becomes less favourable, three outcomes are possible: (1) objective and subjective listening effort will begin to increase at the same SNR, (2) objective effort will increase before subjective effort, or (3) subjective effort will increase before objective effort. Theoretical accounts of listening effort support option (2), with the Model of Listening Engagement proposing that subjective effort is caused by objective effort being elevated to the point where cognitive resources are barely adequate (Herrmann & Johnsrude, 2020a), although no experiment has assessed this proposal. As mentioned, the majority of results are consistent with objective effort being more responsive to cognitive demand than subjective effort, at least under certain conditions (Downs & Crum, 1978; Hicks & Tharpe, 2002; Hornsby, 2013; Pals et al., 2013; Mackersie et al., 2015). These findings imply that the perception of effort may lag behind the exertion of effort. This would be consistent with a broader literature demonstrating that factors beyond conscious awareness can influence ease of perceptual processing. In the visual domain, subliminal prime words can improve the recognition of phonologically similar words (Forster & Davis, 1984; Ferrand & Grainger, 1992). In the speech domain, prior experience with stimuli causes them to be perceived as less noisy when heard again, even if participants are unaware of this experience (Jacoby et al., 1988; Goldinger et al., 1999). In Rodd et al. (2005), participants also listened to sentences that varied in their ambiguity (i.e., the number of words with multiple meanings). High-ambiguity sentences activated a network that included the inferior frontal gyri bilaterally, reflecting greater objective effort. However, only about 10% of participants consciously noticed any ambiguous sentences.

In the current study, I predict that as the SNR becomes less favourable, gist understanding (i.e., speech intelligibility) will remain very high up until the least favourable SNRs. In addition, I predict that as the SNR becomes less favourable, secondary-task reaction time (i.e., objective effort) and effort rating (i.e., subjective effort) will exhibit a generally similar trend, with both measures increasing (Zekveld et al., 2011; Johnson et al., 2015; Lau et al., 2019). These relationships are expected to persist even when accounting for speech intelligibility or other extraneous factors. However, the trends are not expected to be identical for objective and subjective effort. In particular, I predict that objective effort will begin to increase at a more favourable SNR than subjective effort (Downs & Crum, 1978; Hicks & Tharpe, 2002; Hornsby, 2013; Pals et a., 2013; Mackersie et al., 2015), exhibiting greater responsiveness to cognitive demand (Herrmann & Johnsrude, 2020a). For instance, if objective effort first increased from the most favourable conditions at an SNR of +9 dB, then subjective effort would not be expected to increase until an SNR of +8 dB or less favourable. I predict that objective effort will also begin to increase at a more favourable SNR than the one at which speech intelligibility begins to decrease, since cognitive resources recruitment may be able to maintain a high level of speech intelligibility (Houben et al., 2013). These results would support the assumed multidimensionality of listening effort, and they would offer the first experimental support to the proposal that increases in objective effort may precede increases in subjective effort.

Chapter 2

2 Methods

2.1 Participants

Two-hundred twenty-four participants were included in the current study. All were recruited from the Amazon Mechanical Turk online participant pool (https://www.mturk.com) via the participant sourcing platform CloudResearch (Litman et al., 2017). The recruitment pool was narrowed based on three criteria: being located in the United States or Canada, being 18–40 years of age, and speaking English as a first language. These participants consisted of 150 males and 74 females, who ranged in age from 20 to 40 (M = 32.46, SD = 4.84). In terms of their ethnic background, 161 participants identified as being White, 19 identified as Black, 12 as Latin American, 11 as East Asian, and three as Southeast Asian, while 18 identified with more than one ethnic group. Two participants reported less than grade 12 as their highest level of education, 28 had a high school diploma, 75 had some college or university, 105 had a bachelor's degree, two had some postgraduate education, and 12 had a postgraduate degree.

2.2 Materials

2.2.1 Speech stimuli

The stimuli used were from Saleh et al.'s (2020) adaptation of the Connected Speech Test (CST). The original CST (Cox et al., 1987) consisted of 48 test passages, each of which was unified by a single topic (e.g., "Windows") and contained ten (or occasionally nine) English sentences related to that topic (e.g., "Windows provide light and air to rooms"). These passages were masked with six-talker babble. The adapted version of the CST, which includes 32 of the original 48 test passages, was created to correct a number of shortcomings in the original, such as old recording equipment and a fast speaking rate. In the adapted version, the passages were spoken by a single female speaker with a general North American accent. They were recorded at a sampling rate of 44.1 kHz. This version, like the original, also included six-talker babble, which was created from recordings of three male and three female speakers reading five CST non-test passages. The masking babble was unique to each passage. The 32 passages were found to yield equivalent intelligibility in a study of native English-speaking younger adults with normal hearing (Salah et al, 2020). The average passage word report (calculated based on 25 key words per passage) ranged from 32.4% to 67.7% at an SNR of -2 dB.

From here on, all discussion of the CST will refer to the adapted version (Saleh et al., 2020). The current study required only 28 of the 32 CST test passages (see Appendix A). To select the 28 passages, the "Orange" passage was first discarded, for which the end of sentence six was cut off. Of the remaining 31 passages, the three passages with only nine sentences ("Lung", "Lizard", and "Calendar") were designated to be used as practice passages rather than test passages. The 28 remaining test passages were then divided into two lists of 14 passages (see Table 1), List A and List B, such that the mean intelligibility of these lists was identical ($t[26.00] = 0.00$, $p = 1.00$, $d = 0.00$) based on data provided by Saleh et al. (2020).

Praat version 6.1.37 (Boersma, 2001) was used to take the passages, encoded as WAV files, and manually isolate each sentence such that the speech started when the selection started and ended when it ended. This was done to remove the 5 s silences that existed between the sentences in their original Salah et al. (2020) form, allowing for easier processing. Custom scripts in MATLAB version 9.9 (The MathWorks Inc., Natick, MA) were used to modify the sentences further. First, 14 different SNRs were created for each sentence, using the six-talker babble that accompanied each sentence. These conditions included clear (no noise) as well as all SNRs, in 1 dB steps, from -1 dB to +11 dB. Second, the root-mean-square amplitude of each sentence was standardized to 0.10, ensuring that all stimuli (i.e., the speech and noise together) were the same level. Third, all sentences from the same passage were concatenated to form one long clip per passage, now with no silences. The average passages duration was very similar for List A and List B ($t[25.41] = 0.35$, $p = .727$, $d = 0.13$; see Table 1). All passages were then converted to MP3 (constant bit rate = 128 kbps) using Switch version 10.00 (NHC Software, Canberra, Australia) to reduce for their file size for online data collection. Although this may degrade intelligibility somewhat, this would be true across all conditions.

2.2.2 Secondary task

The secondary task used to measure objective effort was a case-judgement task adapted from Rodd et al. (2010). While listening to each passage, participants were presented with a series of letters one at a time. Each letter was selected, in either uppercase or lowercase form, from the following list: ABDEFGHNQRT. These letters were used because their uppercase and lowercase forms were easy to distinguish (e.g., in contrast to “C” and “c”, which are very similar). Letter selection was pseudorandom, with the constraints that the same letter in the same case could not be presented two times in a row, and the same case could not be presented four times in a row.

2.2.3 Effort rating scale

The effort rating scale used to measure subjective effort was adapted from a scale created by Luts et al. (2010) and used in several studies since (Holube et al., 2016; Rennies et al., 2014; Schepker et al., 2016; Holmes et al., 2018). While listening to each passage, participants were continuously presented with a prompt that read as follows: “How much effort do you feel listening to this passage? Click along the number line to make and update your rating.” Participants were also continuously presented with a seven-point visual analog scale with the following labels placed below tick marks from left to right: “No effort”, “Very little effort”, “Little effort”, “Moderate effort”, “Considerable effort”, “Very much effort”, “Extreme effort”. Above these tick marks were corresponding integers from 1 to 7.

2.2.4 Comprehension questions

After each passage, participants were tested on their comprehension. This was included to check that participants paid attention the passages and followed instructions to prioritize listening to them. Comprehension questions were custom-written for this experiment (see Appendix B). Participants were presented with two facts related to the passage topic, both of which were true, but only one of which was stated in the passage. All facts that were stated in the passages were drawn from around the middle of the passage, in particular between sentences four and seven. These stated facts were always rephrased, and occasionally they involved an inference that required remembering more than one sentence. The facts that were not stated in the passage were obtained from outside sources, and they were roughly matched for their length and distinctiveness.

Table 1: Connected Speech Test passage lists. RAU = rationalized arcsine units.

List A			List B		
Passage Topic	**Intelligibility (RAU)**	**Duration (s)**	**Passage Topic**	**Intelligibility (RAU)**	**Duration (s)**
Grass	-8.26	28.45	Glove	-8.05	31.77
Wolf	-8.05	30.98	Donkey	-7.63	31.30
Eye	-7.63	30.15	Eagle	-6.58	32.91
Wheat	-4.47	31.96	Violin	-5.11	32.69
Clock	-4.89	33.65	Woodpecker	-5.32	29.84
Weed	-2.16	35.07	Lawn	-3.63	30.36
Egg	-1.32	28.67	Lead	-1.32	33.84
Ear	-1.11	33.48	Window	-0.89	34.67
Kite	-0.68	30.79	Chimney	-0.05	30.41
Lake	1.84	32.95	Lime	1.00	28.55
Gold	2.89	33.78	Kangaroo	4.16	30.82
Oyster	5.42	30.92	Owl	4.79	30.51
Zipper	6.05	30.87	Knife	6.26	29.56
Lemon	7.11	30.09	Cabbage	7.11	31.12
M	*-1.09*	*31.56*	*M*	*-1.09*	*31.31*
SD	*5.23*	*1.99*	*SD*	*5.19*	*1.71*

2.3 Design

The current study had a within-subject design with one independent variable: SNR (clear, +11 dB, +10 dB, +9 dB, +8 dB, +7 dB, +6 dB, +5 dB, +4 dB, +3 dB, +2 dB, +1 dB, 0 dB, -1 dB). Pilot testing ensured that the passages presented at these SNRs were highly intelligible up until the least favourable SNRs, meaning that motivation and performance were unlikely to bias effort ratings (Wu et al., 2016; Moore & Picou, 2018). The study consisted of two blocks: an objective block and a subjective block. In the objective block, participants listened to 14 speech passages (one at each SNR) while their reaction times were measured on the secondary case-judgement task. In the subjective block, they listened to a different 14 speech passages (one at each of the same SNRs) while rating listening effort subjectively in real time. This two-block design ensured that subjective ratings were not obscured or biased by a secondary task (Yeh & Wickens, 1988; Recarte et al., 2008). After each passage, participants' gist understanding (yes or no) and passage comprehension (correct or incorrect) were also measured.

Fifty-six counterbalancing conditions were used, each completed by four participants. To produce this number of conditions, three features were counterbalanced across participants: the order of the two blocks (two options), which passage list—List A or List B—was assigned to each block (two options), and how the SNRs were assigned to the 14 passages within each block (14 options). In the latter case, one participant may have had a condition in which passage 1 assigned to clear, passage 2 to +11 dB, passage 3 to +10 dB, and so on. In the next condition, passage 1 was instead assigned to +11 dB, passage 2 to +10 dB, passage 3 to +9 dB, and so on. This pattern continued across the 14 options and was used for both blocks. Across participants, each passage was presented at each SNR once. Within each block, the passages (and thus SNRs) were presented in a random order for each participant. The study was approved by Western University's Non-Medical Research Ethics Board (Project ID 119968; see Appendix C).

2.4 Procedure

Participants were encouraged to complete the current study using Google Chrome or Firefox. The study began with participants being directed to a Qualtrics link to complete a questionnaire (Qualtrics, Provo, UT). After providing informed consent through a checkbox, they completed a demographic questionnaire. Participants then completed an auditory calibration. The calibration required participants to listen to a CST practice passage ("Calendar", +11 dB), set to the same root-mean-square amplitude as the later speech clips, and adjust their computer volume to a comfortable listening level. They were told that they would have to maintain their volume at this level for the rest of the experiment. Participants were then redirected to Pavlovia (https://pavlovia.org), which hosted an experiment programmed using PsychoPy version 2021.2.3 (Peirce et al., 2019). Pavlovia-hosted PsychoPy experiments have been shown to demonstrate extremely high reaction time precision (within 3.5 ms) across all common browsers (Bridges et al., 2020). During this redirection, each participant's unique alphanumeric Qualtrics ID was passed to Pavlovia, and they were also assigned a unique sequential numeric ID using a webpage-based tool (Morys-Carter, 2021).

This phase of the study began with participants completing a headphone check. The check required them to listen to sets of three tones: two diotic white noise stimuli and one dichotic stimulus that evokes the Huggins pitch, an illusory pitch phenomenon (Cramer & Huggins, 1958). To become familiar with the task, participants first heard a series of three sets of these tones back-to-back, and they were told that the second tone in each set contained a "beep" (i.e., the Huggins pitch). They were then presented with six sets of tones, which, for each participant, were randomly sampled without replacement from a list of 18 sets (six with the correct response in position 1, six in position two, and six in position three). Participants indicated which tone contained the Huggins pitch by pressing their <1>, <2>, or <3> key. As described by Milne et al. (2020), this check was designed to be easy for participants using headphones and very hard for those using speakers. Participants passed the check if they responded correctly for at least five sets. A cut-off of five is slightly more lenient than Milne et al.'s recommendation that all six be correct. This increased the estimated rate at which headphone users could be correctly detected (85% vs. 80%), at the expense of passing more users of speakers (30% vs. 20%; Milne et al., 2020). Given that other measures were put in place to detect the users of speakers (e.g., asking about their sound presentation method in the debrief), this was considered a practical trade-off. Participants then read about the structure of the experiment to come. To familiarize themselves with these passages and their range of difficulty, participants were presented with a practice passage at a more favourable SNR than they would listen to during the experiment (always "Lizard", +22 dB) followed by one at practice passage at a slightly les favourable SNR than they would listen to (always "Lung", -2 dB).

Participants then completed the experiment proper. Before the first block (e.g., objective), participants were given thorough instructions on the tasks to come, including the gist and comprehension questions. They next completed a practice passage for the objective block (always "Lung", +11 dB). When ready, participants then proceeded to the 14 test passages. Each passage was introduced by a 3-s countdown. After the countdown, a black box with a white border appeared centred on the screen. This was accompanied by prompts on either side of the box instructing participants, once the letters begin to appear, to press <Q> for "Uppercase" and <P> for "Lowercase". This box and prompt were displayed for 2 s before the passage was played. Once the passage began, one letter was presented per sentence. The letters had a stimulus duration of 1500 ms, longer than the 1000 ms used by Rodd et al. (2010). They were presented with a standard inter-onset interval for each passage, chosen such that, on average, a letter was presented 250 ms before the midpoint of each sentence. This ensured that most letters were presented within only one sentence, and that the final letter of each passage could always be presented in full before the passage concluded. This contrasts the more precise approach of Rodd et al. (2010), which presented letters at the offset of the second homophone in each sentence. Given that there was no hypothesis as to when listening effort would be greatest, such precision was not necessary. If participants did not respond within the 1500 ms stimulus duration, the letter disappeared and responses could no longer be registered. See Appendix D for the instructions and prompt used for the secondary task.

After the passage had concluded, participants were asked "Did you understand the gist of the passage?", pressing <Q> for "Yes" and <P> for "No". This was used as an index of speech intelligibility. Listeners' subjective ratings of speech intelligibility generally agree with objective measures (Cox et al., 1991; Larsby & Arlinger, 1994; Cienkowski & Speaks, 2000), and gist understanding in particular has been found to agree with word report scores (Wild et al., 2012). Participants were then presented with two true statements about the passage topic and asked "Which of these true facts was stated in the passage?", pressing <Q> for the top option and <P> for the bottom option. Participants were encouraged to guess if they were unsure of any of the answers. Half of the comprehension questions within each list had the top option as the correct answer, while for the other half it was the bottom option. The location of the correct answer was fixed for each passage (e.g., the "Windows" passage always had the correct answer on the bottom). After responding to a question, another countdown began, introducing the next passage. After seven passages (halfway through the block), participants had a chance to take a break before continuing on when ready. After all 14 passages of the first block, they were given another chance to take a break before continuing onto the second block.

Before the second block (e.g., subjective), participants were once again given thorough task instructions. These instructions included being introduced to the listening effort scale, which they were given a chance to use and become familiar with. Participants then completed a practice passage for the subjective block (always “Lizard”, +11 dB). In the second block, the instructions and practice did not include the gist and comprehension questions, since these were already explained in the first block. Each test passage was again introduced by a 3-s countdown, after which the listening effort scale and associated prompt appeared on the screen 2 s before the passage began playing. Participants were instructed to click along the scale to indicate an effort rating, which placed a circular red marker. The wording of the prompt and instructions emphasized listening effort as a feeling rather than a process. The instructions also emphasized the distinction between effort and perceived performance or difficulty, following the recommendation of Moore and Picou (2018) to minimize response bias. Participants were also instructed to update their rating whenever they felt effort changing. The scale and prompt for a passage would only disappear if the participant had made at least one effort rating and the passage had concluded, plus an additional 2 s to allow in-progress ratings to be registered. In this block, after each passage, participants once again answered gist and comprehension questions. They were again able to take a break after seven sentences and at the end of the block. See Appendix E for the instructions and prompt used for the effort rating scale.

Once participants completed both blocks, they were directed to another Qualtrics link to complete a second questionnaire. Their unique alphanumeric Qualtrics ID from the first questionnaire was again passed with them to Qualtrics. Within this questionnaire, participants began by completing a debrief, in which they were asked a series of questions about their performance during the preceding listening tasks. This included an open-ended question where participants could report any relevant details about their experience. Finally, they completed a short-form version of the Raven’s Advanced Progressive Matrices (RAPM), used to measure fluid intelligence (Arthur & Day, 1994). In this task, participants were presented with a series of three-by-three grids of shapes, with the bottom-right shape missing. From six options, they had to select the shape that belongs in the missing space by preserving the relationship among the other shapes. Participants had 20 mins to solve all 12 items, with a countdown on the top-left of the screen displaying their time remaining. They were encouraged to guess if they were unsure of any of the answers. This task scored participants' responses and assigned them a score from 0 to 12. Once participants were finished the task (or ran out of time), they were given a completion code to copy and submit to MTurk and CloudResearch to receive their compensation, equivalent to $10 USD. At this point, they were then free to exit their browser window.

2.5 Data quality

Online research generally replicates the findings of in-person research (Gosling et al., 2004; Buhrmester et al., 2011; Mason & Suri, 2012; Berinsky et al., 2014; Thomas & Clifford, 2017; Buchanan & Scofield, 2018). Further, my lab has previously reported consistent results between online Pavlovia-hosted experiments and in-person experiments using a dual-task listening paradigm (Bain et al., 2020). Nonetheless, at the recruitment stage of the current study, a number of CloudResearch features were used to optimize participant quality. Importantly, only CloudResearch approved participants were used, which is a select group of over 50,000 vetted, high-quality participants who have passed a series of attention and engagement measures. In addition, duplicate IP addresses were blocked, as were participants from suspicious geocode locations and those whose IP addresses did not match their stated locations. Using CloudResearch with such criteria has been found to yield high participant attentiveness, honesty, comprehension, reliability, as well as overall data quality (Eyal et al., 2021).

A further 157 participants completed the study but were excluded based on pre-determined criteria (with the exception of the late effort-rating criterion; see below). These included self-reporting any of the following: having a hearing impairment (n = 3), having an uncorrected vision impairment (n = 11), not completing the listening tasks to the best of their abilities (n = 2), not completing the listening tasks in a quiet environment (n = 0), not using a computer (n = 1), or not using headphones (n = 16). Participants were also excluded if they failed the check used to confirm headphone use (n = 95), or if their average passage comprehension was below chance for either the subjective block (n = 6) or objective block (n = 4). For many of the remaining trials, participants did not rate effort until after the passage had concluded. The large majority of participants (90.53%) did this for three or fewer of their 14 effort ratings, while most other participants (7.00%) did for 12 or more of their ratings. Since such late responding likely reflected either inattention or a failure to understand task instructions, participants were excluded who did this for more than half of their 14 passages (n = 19).

2.6 Data analysis

2.6.1 Pre-processing

In the objective block, participants made up to ten case-judgments per passage, yielding up to ten reaction times. Prior to processing, 1.12% of all reaction time data were missing. This included cases where a technical error caused a participant's reaction time data not to be saved for a particular letter position (0.27%) as well as cases where a participant failed to respond within 1500 ms (0.85%). First, all reaction times corresponding to incorrect trials (e.g., selecting "P" for lowercase when the stimulus was "A") were excluded from further analysis (2.17% of trials). Second, reaction times less than 100 ms were excluded (only one trial), since these are likely too fast to reflect genuine responding (Whelan, 2008). Third, the reaction time data were normalized using a natural log transformation (Whelan, 2008), and the average reaction time was calculated for each case-judgement letter position (1–10) within each participant. Pairwise t-tests with Holm-Bonferroni correction (Holm, 1979) were used to explore the effect of letter position on reaction times. These revealed that reaction times in the first letter position were significantly slower than all other letter positions (ps < .001), with no other letter positions differing from one another (ps > .097; see Appendix F). This likely reflects the unexpectedness of the first letter being presented. As a result, reaction times from the first letter position were excluded from further analyses. Fourth, log-transformed reaction times that were more than three standard deviations from the mean of their SNR across all participants were excluded (1.07% of trials for letter positions 2–

10), which removed outliers while minimizing bias and data loss (Berger & Kiefer, 2021). Fifth and finally, log-transformed reaction time for letter positions 2–10 were averaged separately for dominant-hand and non-dominant-hand responding, yielding two values per participant within each SNR (Rodd et al., 2010). After all processing, averaged reaction time (separated by hand-dominance) was missing from 0.67% of passages.

In the subjective block, participants rated effort continuously within each passage. Prior to processing, no effort ratings were missing, since participants were required to rate effort to move onto the next passage. These continuous effort ratings were converted to a single time-weighted average rating for each passage. If a participant made only one effort rating in a passage, this was considered their average effort rating. However, if this singular rating came after the end of the passage (2.14% of passages), it was removed from the analysis. If a participant made multiple effort ratings, their first rating was assumed to reflect their effort from the start of the passage up to the second rating, any intermediate ratings were assumed to reflect their effort from the time it was made up to the next rating, and their final rating was assumed to reflect their effort from the time it was made up to the end of the passage. The amount of time spent at each effort rating was used to calculate time-weighted average effort ratings. Effort ratings that were made in the 2 s before or after the passage were excluded from this average. Finally, as in the objective block, average effort ratings more than three standard deviations from the mean of their SNR across all participants were excluded (0.64% of passages). After all processing, average effort ratings were missing for 2.77% of passages.

In both blocks, participants rated gist understanding and answered a comprehension question after each passage. For gist understanding, answering "no" (i.e., they did not understand the gist) was scored as 0, while answering "yes" was scored as 1. For comprehension, incorrect responses were scored as 0, while correct responses were scored as 1. Prior to processing, no gist or comprehension data were missing, since participants were required to answer these questions to move on. However, some understanding data were excluded based on listening effort data. In particular, the corresponding understanding scores were excluded in cases where the only effort rating came after the end of the passage, since this likely reflects a lack of attention. This resulted in gist and comprehension scores being missing for 2.14% of passages.

2.6.2 Statistical analysis

All data were analyzed using R version 4.1.1 (R Core Team, 2021). Mixed-effects modelling, implemented using the package "lme4" version 1.1.27.1 (Bates et al., 2015), was used to model the effect of SNR on all dependent variables. For secondary-task reaction time and effort rating, the clear condition was subtracted from the other SNRs as a baseline (Wu et al., 2016). For all analyses, the clear condition was not included unless otherwise stated. Models were fit using restricted maximum likelihood estimation (or maximum likelihood estimation during model comparison; Gurka, 2006) and the bound optimization by quadratic approximation algorithm (Powell, 2009), with the maximum number of function evaluations set to 2×10^5. A crossed-random design was used in which intercepts were allowed to vary randomly by participant and (when possible) by passage (Carson & Beeson, 2013). The procedure for selecting random slopes largely followed that of Field et al. (2012). Random slopes with respect to participant were added for each SNR term sequentially, but they were dropped if a likelihood ratio test did not reveal a significantly improved model fit or if the produced model had a singular fit. Random slopes were not added for covariates, or with respect to passage, since there was no expectation that passages would be differently affected by SNR.

Models of all dependent variables included linear SNR (-1–11), negative exponential SNR (e^{-SNR}), and RAPM score. Negative exponential SNR was included to account for the expected rapid changes in speech understanding and listening effort that may occur as the SNR became less favourable, while RAPM score was included as a covariate to account for differences in speech understanding or listening effort that may be related to intelligence (Michalek et al., 2018). In addition to these factors, reaction time and effort rating models included passage-level gist understanding as a covariate to ensure that differences in listening effort between SNRs are not driven by speech intelligibility (Moore & Picou, 2018; Winn & Teece, 2020). Reaction time models also included button-press hand-dominance as a covariate (0=non-dominant, 1=dominant) to account for faster responding with the dominant hand (Rodd et al., 2010). As a result, reaction time analyses used data averaged separately for non-dominant and dominant responses within each passage, while effort ratings only had one average per passage. No interactions were predicted or included in these models.

Mixed-effects model parameters were obtained from model summaries and significance was assessed using “lmerTest” version 3.1.3 (Kuznetsova et al., 2017) with the Satterthwaite approximation of degrees of freedom (Satterthwaite, 1946), as recommended by Luke et al. (2017). For secondary-task reaction time, a model of raw reaction times—built identically to the analyzed log-transformed model—was used to obtain interpretable unstandardized regression coefficients. For all dependent variables, standardized regression coefficients (β) were calculated using “MuMIn” version 1.43.17. In cases where an effect of SNR was significant, another model was fit with linear SNR (here including the clear condition) coded as a factor, but with no non-linear terms or random slopes. Pairwise comparisons were made between estimated marginal means for all 91 SNR pairings, with Holm-Bonferroni correction, using “emmeans” version 1.7.0. For each final model, the coefficient of determination (R^2) was also calculated based on the approach of Nakagawa and Schielzeth (2012) using the package “MuMIn”.

To compare the responses of gist understanding, secondary-task reaction time, and effort rating to SNR, the best equation was determined for each of these using the “mosaic” package version 1.8.3. This package used as input one data point per SNR, averaged across participants. For reaction time, this average was calculated using passage reaction times that were averaged across all nine letters (2–10) and not separated by button-press hand-dominance. The forms of the equations (e.g., linear vs. negative exponential) were chosen based on the mixed-effects modelling results: any significant term was included in the equation. Adjusted R^2, which penalizes model complexity, was also calculated for each equation. In cases where negative exponential terms were used, the “knee point” of the equation (i.e., the point of maximum curvature) was determined with the kneedle algorithm (Satopää et al., 2011) using the package “kneedle” version 1.0.0 (default parameters). This represented an estimate of the SNR after which listening effort began to change rapidly as the SNR became less favourable.

Chapter 3

3 Results

If participants followed instructions to prioritize listening to the passages, then their understanding of the passages should have been similar across the subjective and objective blocks. Comparing participants' gist understanding in each block (averaged across passages) revealed a significant difference between blocks (t[223] = 2.68, p = .008, d = 0.18), with gist understanding 1.25 percentage-points higher in the objective block. However, this effect was very small and went in the opposite direction to what would be expected if participants had failed to prioritize the primary task during the objective block. This same comparison for comprehension revealed no significant difference (t[223] = -0.97, p = 0.332, d = 0.07), with comprehension only 0.71 percentage-points higher in the subjective block. These results indicate that there was little to no difference in understanding between the two blocks. As a result, comprehension and gist understanding were averaged across the two blocks before analysis. For mixed-effects modelling of comprehension, see Appendix G. This offers confidence that participants followed instructions to prioritize passage listening even in the objective block, meeting this criterion for the interpretation of secondary-task reaction time as an index of objective effort.

3.1 Gist understanding

Figure 1 shows the effect of SNR on gist understanding. Average gist understanding was over 95% in the most favourable SNRs up until the +1 dB SNR, at which it declined rapidly down to approximately 75% in the least favourable SNR. The final model of gist understanding included random intercepts for participants and random slopes for negative exponential SNR with respect to participant. The effect of linear SNR on gist understanding was significant (b = 0.38, 95% CI = [0.20, 0.57], β = 0.08, t[222.13] = 4.02, p < .001). The effect of negative exponential SNR was also significant (b = -6.92, 95% CI = [-8.53, -5.30], β = -0.28, t[206.93] = -8.41, p < .001). Comparisons between SNRs (including clear) revealed that gist understanding was lower at +1 dB than all

SNRs from the clear condition to +4 dB (*p*s < .031), lower at 0 dB than all more favourable SNRs (*p*s < .001), and lower at -1 dB than all more favourable SNRs (*p*s < .001). The effect of RAPM score was also significant (b = 0.55, 95% CI = [0.13, 0.98], β = 0.08, t[225.65] = 2.58, p = .01), with each one-point increase in RAPM score associated with an increase in gist understanding of just over 0.5 percentage-points on average (see Appendix H for the distribution of RAPM score). The conditional R^2 of the final model was .75, meaning that the fixed and random effects accounted for 75% of the variance in gist understanding.

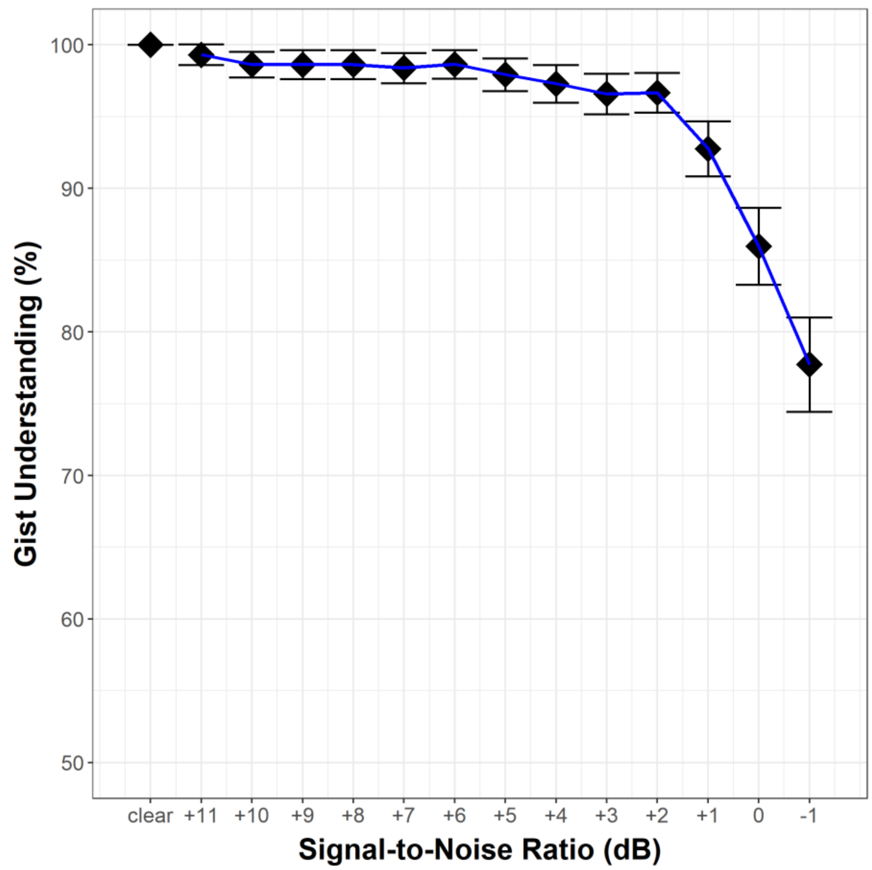

Figure 1: Gist understanding by SNR. Gist understanding served as an index of speech intelligibility. Data points represent the means of each condition, error bars represent the 95% confidence intervals around the means, and the blue lines connecting the means show the trend. The clear condition is excluded from this line since it was categorically different from the other SNRs.

3.2 Secondary-task reaction time

Figure 2 shows the effect of SNR on secondary-task reaction time. For context, average secondary-task accuracy (not shown or analyzed) was approximately 97%, indicating engagement with the task. Average reaction time was between 560 ms and 575 ms in the more favourable SNRs, then increasing sharply in the less favourable SNRs up to more than 585 ms in the least favourable SNR. The final model of reaction time included random intercepts for participant and passage but no random slopes. The effect of negative exponential SNR on reaction time was significant (b = 6.08, 95% CI = [2.50, 9.68], β = 0.05, t[5526] = 3.29, p < .001), but not the effect of linear SNR (b = -0.35, 95% CI = [-1.05, 0.34], β = -0.01, t[5511] = -0.99, p = .321). Comparisons between SNRs (including clear) revealed that reaction time was slower at -1 dB than at +5 dB, +6 dB, +8 dB, +9 dB, +11 dB, and clear (ps < .029). The effect of button-press hand-dominance was also significant (b = -3.93, 95% CI = [-7.91, 0.04], β = -0.02, t[5521] = -2.12, p = .034), with responses just under 4 ms faster when made with the dominant hand. The effect of RAPM score was not significant (b = 1.13, 95% CI = [-1.80, 4.05], β = 0.02, t[222.5] = 0.60, p = .551; see Appendix H) nor was the effect of gist understanding (b = 0.04, 95% CI = [-0.08, 0.17], β = 0.01, t[5738] = 0.90, p = .370). The conditional R^2 of the final model was .37, meaning that the fixed and random effects accounted for 37% of the variance in reaction time.

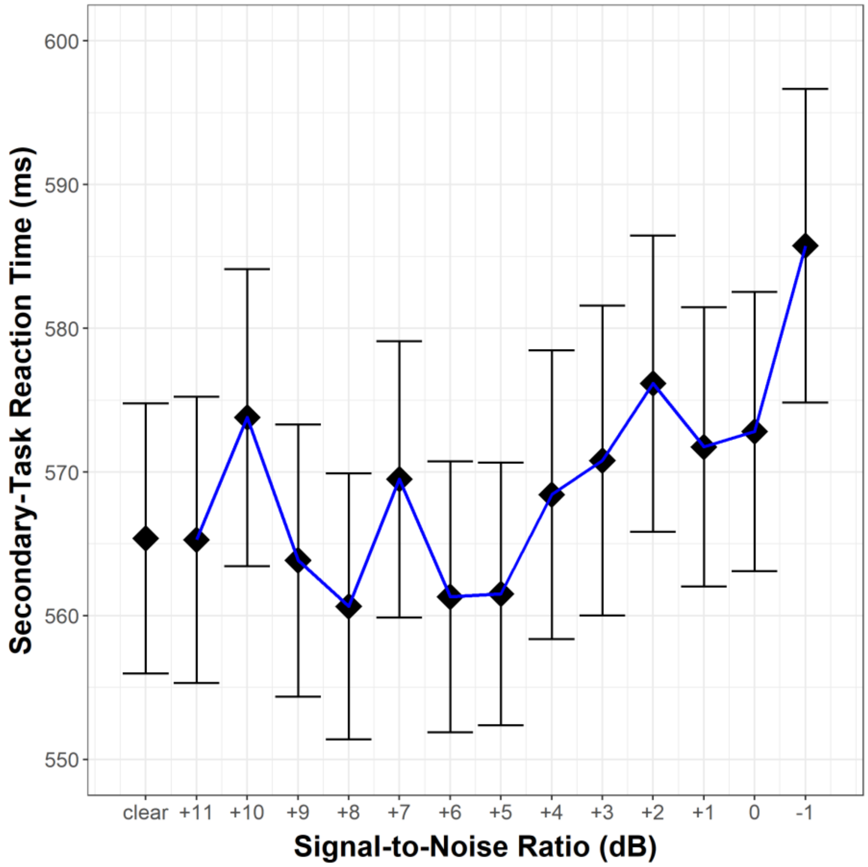

Figure 2: Secondary-task reaction time by SNR. The secondary task was a case-judgement task, and reaction time on this task served as an index of objective effort. Data points represent the condition means, error bars represent the 95% confidence intervals, and blue lines connecting the means to show the trend.

3.3 Effort rating

Figure 3 shows the effect of SNR on effort rating. Average effort rating was approximately 3/7 at the most favourable SNR and then increased steadily as the SNR became less favourable, ending at approximately 6/7 at the least favourable SNR. The final model of effort rating included random intercepts for participant and passage as well as random slopes for linear SNR with respect to participant. The effect of linear SNR on effort rating was significant (*b* = -0.24, 95% CI = [-0.66, 0.02], *β* = -0.26, *t*[323] = -39.57, *p* < .001), but not the effect of negative exponential SNR (*b* = 0.03, 95% CI = [0.02, 0.01], *β* = -0.01, *t*[2350] = 1.49, *p* = .137). Comparisons revealed that effort rating differed between all SNRs (including clear), with the less favourable SNR always associated with a higher effort rating (*p*s < .014). The effect of gist understanding (scaled to 10 percentage-points) was significant (b = 0.04, 95% CI = [-0.05, 0.02], β = 0.00, t[2483] = -4.76, p < .001), with each 10 percentage-point increase in gist understanding associated with a decreased in effort rating of approximately 0.04 on average (see Appendix H). The effect of RAPM score was not significant (b = -0.01, 95% CI = [-0.04, 0.04], β = -0.05, t[214.2] = -0.70, p = .487). The conditional R^2 of the final model was .80, meaning that the fixed and random effects accounted for 80% of the variance in effort rating.

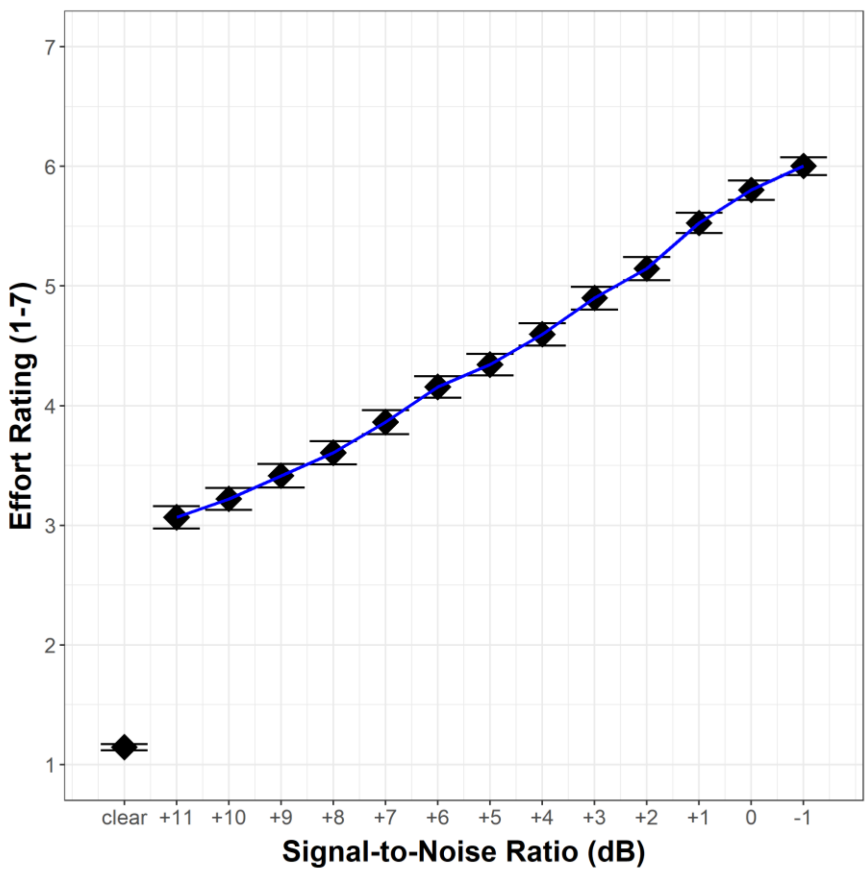

Figure 3: Effort rating by SNR. Effort rating served as an index of subjective effort. Data points represent the condition means, error bars represent the 95% confidence intervals, and the blue line connects the means to show the trend.

3.4 Comparing measures

Figure 4 shows the functions that best model gist understanding (A), secondary-task reaction time (B), and effort rating (C). As the SNR became less favourable, effort ratings appeared to increase first, followed by secondary-task reaction time. There also appears to be a narrow range of SNRs at which reaction time is elevated but gist understanding remains near ceiling. Mixed-effects modelling revealed that linear and negative exponential SNR both predicted gist understanding. The best equation for gist understanding was as follows:

$$\text{gist understanding} = 0.02(\text{SNR}) + -e^{-0.61(\text{SNR})+2.44} + 98.62 \quad (1)$$

The adjusted R^2 of this model was .99, accounting for 99% of the variance in average gist understanding. The knee point was where the SNR was +2 dB. Before this point (between +11 dB and +2 dB), gist understanding was relatively flat, while after this point, gist understanding decreased sharply. Mixed-effects modelling revealed that negative exponential SNR predicted secondary-task reaction time. The best equation for log-transformed reaction time was as follows:

$$\text{reaction time} = e^{-0.44(\text{SNR})-3.98} +0.00 \quad (2)$$

The adjusted R^2 of this model was .55, accounting for 55% of the variance in reaction time. The knee point was where the SNR was +3 dB. Up to this point (between +11 dB and +3 dB), reaction time was relatively flat, while after this point, reaction time increased sharply. Mixed-effects modelling revealed that linear SNR predicted effort rating. The best equation for effort rating was as follows:

$$\text{effort rating} = -0.26(\text{SNR}) + 4.56 \quad (3)$$

The adjusted R^2 of this model was .99, accounting for 99% of the variance in effort rating. Effort rating increased linearly across all SNRs, from +11 dB to -1 dB. In sum, as the SNR became less favourable, effort ratings increased first (even at the most favourable SNRs), then reaction time appeared to increase next (SNR > +3 dB), before gist understanding finally appeared to decrease at the least favourable SNR of these three measures (SNR > +2 dB).

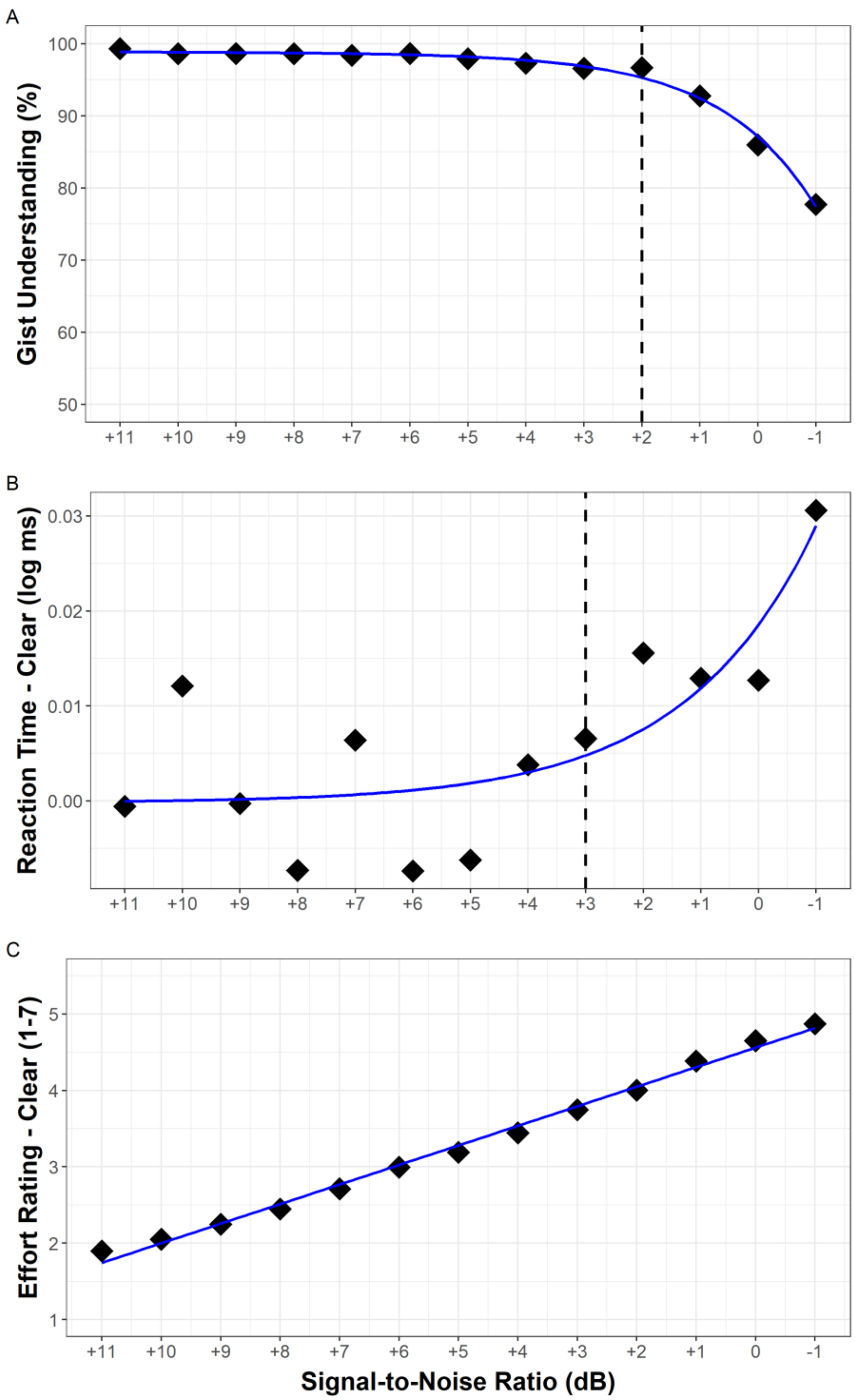

Figure 4: Functions for all measures. Data points represent the condition means; solid blue lines represent the equations that best describe gist understanding (A),

secondary-task reaction time (B), and effort rating (C); and the dashed black lines represent the points of maximum curvature for non-linear functions.

Chapter 4

4 Discussion

The current study was motivated by the lack of research comparing objective and subjective measures of listening effort. In particular, despite the theoretical proposal that objective effort may be more responsive to cognitive demand than subjective effort (Herrmann & Johnsrude, 2020a), no experiment had compared their response profiles directly. Thus, it has been unclear whether, as cognitive demand increases, one dimension of effort would begin to increase before the other. The results indicated that gist understanding was very high across the majority of SNRs. As the SNR became less favourable, effort ratings increased before any other measure, showing a response to SNR changes even when the SNR was most favourable. When the SNR became less favourable than approximately +3 dB, secondary-task reaction time began to increase as well, doing so at a very rapid rate. Finally, only after an SNR of approximately +2 dB did gist understanding begin to rapidly decrease.

4.1 Objective effort increased as the SNR became less favourable

As I predicted, secondary-task reaction time (i.e., objective effort) eventually increased as the SNR became less favourable. The size of this increase (≈ 20 ms between clear and -1 dB) is in line with prior dual-task studies of listening effort (Rodd et al., 2010; Picou et al., 2013). This increase in objective effort is consistent with previous single-measure studies reporting that objective effort increased along with background noise (single-task: Houben et al., 2013; dual-task: Brown & Strand, 2019; peripheral: Kramer et al., 1997; neural: Du et al., 2014). For instance, in Kramer et al. (1997), participants with and without hearing loss listened to speech presented in various levels of fluctuating noise, with the pupil response greater to less favourable SNRs. The eventual increase in objective effort likely reflects the recruitment of additional cognitive resources to compensate for the greater demands. As mentioned, competing talkers introduce energetic and informational masking. For energetic masking, cognitive mechanisms must have been used to “fill in” the occluded speech (Johnsrude & Rodd, 2016). If this occlusion is brief (e.g., a vowel), early-stage auditory processing can automatically repair the speech so that it is heard as a complete whole (Ciocca & Bregman, 1987; Heinrich et al., 2008). If the occlusion is greater (e.g., a word), more deliberate processing may be necessary, with context as well as linguistic and general world knowledge used to resolve perceptual ambiguities (Bashford et al., 1992; Shahin et al., 2009). For informational masking, cognitive mechanisms must have been used to perceptually segregate the competing streams (Van Hedger & Johnsrude, 2021). This is facilitated by higher-level cognitive functions including verbal working memory and selective attention (Brungart, 2001; Humes et al., 2006; Zekveld et al., 2013). These effort-related functions are associated with activation of the anterior portion of the ventral speech pathways, including the inferior frontal gyrus (Wild et al., 2012; Ritz et al., 2022).

Objective effort did not begin to increase at the most favourable SNRs; rather, it only increased once an SNR less favourable than approximately +3 dB was reached. This indicates that, at the most favourable SNRs, speech processing was still relatively automatic, with any increase in cognitive resource recruitment being negligible (Peelle, 2018). This may relate to the specific resources recruited. Early stages of speech processing are relatively automatic, including processing in early auditory cortical regions. However, higher up in the auditory pathway, processing involves effortful re-analysis of speech (Shtyrov, 2010; Wild et al., 2012; Ritz et al., 2022). Processing may have become more effortful as the level of the babble masker approached and exceeded that of the target speech. This would have caused informational masking to become more prominent (Ihlefeld & Shinn-Cunningham, 2008), since participants would have a harder time using differences in sound level to perceptually segregate the target speech from the noise (van Noorden, 1975). This demand likely required effortful processing, such as inhibiting or selectively attending to streams (Nusbaum & Schwab, 1986). These results are consistent with another dual-task study, Brown and Strand (2019), in which normal-hearing younger adults listened to words presented in steady-state speech-shaped noise across a wide range of SNRs. At the same time, participants completed a secondary parity-judgement task in which they classified numbers as either even or odd. Secondary-task reaction times were relatively stable from +14 dB to +2 dB, only beginning to increase at less favourable SNRs. As in the current study, this may reflect the greatly increased difficulty of processing speech with noise at a similar level to the target speech.

Another explanation for the late increase in objective effort is that the dual-task paradigm used was not sufficiently sensitive. Across studies using different secondary tasks, results have been found to vary. For instance, Picou et al. (2013) used a simple visual probe task to find that visual cues did not affect reaction time, while Fraser et al. (2010) used a more complex secondary task and found slower reaction times when visual cues were present (but see Pals et al., 2013; Wu et al., 2014, 2016). In light of these findings, Picou and Ricketts (2014) investigated whether the secondary task used can affect sensitivity to objective effort. In a first experiment, adults with normal hearing listened to words presented with or without babble noise (as well as with or without visual cues). As they listened, participants also completed one of three secondary tasks: a simple task in which they responded to a visual probe, a complex parity-judgement task, and a semantic task in which they judged whether or not words were nouns. In the noise condition, secondary-task reaction time was greater only for the semantic task, which was the most complex and required the deepest processing of the three (Craik & Lockhart, 1972; Eysenck & Eysenck, 1979). In contrast, the simple and complex secondary tasks did not demonstrate greater reaction times in the noise condition. This suggests that secondary tasks with a high cognitive requirement may be the most sensitive to objective effort.

Picou and Ricketts (2014) argued that simpler secondary tasks may be less sensitive because they can be done relatively automatically, meaning that they do not compete with the primary task for cognitive resources (Hasher & Zacks, 1979). In addition, deep-processing tasks may increase the cognitive overlap between the primary and secondary tasks by recruiting more domain-general, rather than modality-specific, cognitive resources (Baddeley & Hitch, 1974; McLeod, 1977). This may be especially relevant when the primary and secondary tasks are of different sensory modalities (e.g., listening to speech and responding to visual stimuli). However, I propose a more precise explanation for their findings. As described by Gagné et al. (2017), for a dual-task paradigm to function as intended, all cognitive resources must be recruited by the combination of the primary and secondary task. This ensures that, if the primary task becomes more challenging, it will be reflected in reduced primary or secondary task performance. Instead, if not all resources are occupied, participants will have cognitive capacity leftover. This leftover capacity could be used to compensate for the increased challenge in the primary task without performance consequences, rendering any increase in objective effort effectively invisible to researchers. Thus, in Picou and Ricketts (2014), it is possible that the simple and complex secondary tasks may not have been automatic, but rather easy enough to leave some resources uncommitted. This is also consistent with their second experiment, which found that even the simple and complex secondary tasks were sensitive to noise for older adults. This is likely because, with older adults having a lower cognitive capacity (Mattay et al., 2006), even simple secondary tasks were enough to ensure that all resources were committed.

Based on this interpretation of Picou and Ricketts (2014), the secondary case-judgement task used in the current study could have been too simple for use in normal-hearing younger adults, rendering it relatively insensitive to objective effort. It is possible that in the most favourable SNRs, not all cognitive resources were occupied by the primary and secondary task. As the SNR became less favourable, previously uncommitted resources could therefore be used to compensate for this greater demand without behavioural consequences. This would make it seem as though objective effort was not initially increasing as the SNR became less favourable, even if more cognitive resources were indeed being recruited. As a result, it is possible that the genuine increase in objective effort occurred at an SNR more favourable than +3 dB, with +3 dB instead representing the SNR after which participants' cognitive capacities were finally occupied in full. This alternative explanation is especially plausible in a study such as this one, in which gist understanding (i.e., speech intelligibility) was near ceiling across the majority of SNRs. This is because the primary listening task was very unlikely to occupy all cognitive resources on its own, meaning that the secondary task would be required to occupy all remaining resources. In contrast, for studies with intelligibility below ceiling, all cognitive resources could presumably be dedicated to the primary task alone. This makes it far less likely that the secondary task would be insufficiently complex.

Despite these concerns, there is some reason to believe that the secondary task was appropriate in the current study. For instance, although intelligibility was at ceiling, participants were not simply tasked with understanding the words, but rather understanding the meaning of the passages. In addition, the secondary case-judgement task was initially developed and used in Rodd et al. (2010). Given that no noise was present, the speech in Rodd et al. (2010) was fully intelligible to younger and middle-aged adults. Nonetheless, the case-judgement task was sensitive to changes in objective effort resulting from ambiguity. In addition, Brown et al. (2020) recently used a similar parity-judgement task to measure objective effort in normal-hearing younger adults. Once again, even with speech being fully intelligible, the secondary task was able to detect greater objective effort when listening to speech with an unfamiliar nonnative accent. However, the current study also differs from these two prior studies in key ways. In Rodd et al. (2010), the cognitive demands of the case-judgement task (selecting the appropriate response) overlapped with those of ambiguous speech (selecting the appropriate meaning), allowing them to compete for resources. In the current study, speech was not ambiguous and was instead masked by noise. This may have reduced the resource overlap, especially at the more favourable SNRs, which would have had less energetic masking and therefore less ambiguity. Further, in Brown et al. (2020), the parity-judgement task used was technically semantic in nature, whereas the case-judgment task used in the current study was orthographic (Fias, 2001). As a result, the case-judgement task may have been less demanding and therefore less sensitive to objective effort.

In the current study, as the SNR became less favourable, objective effort increased (after a +3 dB SNR) before speech intelligibility decreased (after a +2 dB SNR). However, this comparison should be interpreted cautiously given the noise present in the secondary-task reaction time data. Listening effort has often been thought to track intelligibility, with effort increasing only when intelligibility decreases (Zekveld et al., 2010). In recent years, though, the dissociation between listening effort and intelligibility has been acknowledged (Wild et al., 2012; Winn & Teece, 2020; Ritz et al., 2022). Indeed, many studies have reported objective effort changing while intelligibility remained the same. This may reflect the recruitment of additional cognitive resources to maintain high levels of intelligibility for as long as possible (Strauss & Francis, 2017; Peelle, 2018). For instance, in Houben et al. (2013), normal-hearing adults listened to digit triplets presented in stationary noise across a range of SNRs. At the same time, reaction time in a single-task paradigm served as a measure of objective effort. As the SNR changed from +4 dB to -1 dB, objective effort increased, while intelligibility remained at ceiling. Intelligibility did not begin to suffer until an even less favourable SNR of -6 dB was reached. Such compensation has also been demonstrated in clinical populations. In Gatehouse and Gordon (1990), when participants with hearing loss were not wearing their hearing aids, secondary-task reaction time increased despite speech remaining fully intelligible. Ceiling performance is not necessarily required to achieve this effort-intelligibility dissociation. A number of studies have reported that hearing aid parameters can affect secondary-task reaction time even when intelligibility has room to improve (Baer et al., 1993; Sarampalis et al., 2009; Desjardins & Doherty, 2014).

4.2 Subjective effort responded before objective effort

Contrary to my predictions, effort ratings (i.e., subjective effort) increased in response to SNR even in the most favourable conditions, despite objective effort not increasing until an SNR of less favourable than +3 dB was reached. Although this result is consistent with the assumed multidimensionality of listening effort, it appears to disagree with the Model of Listening Engagement (Herrmann & Johnsrude, 2020a), which proposed that objective effort should be more responsive to cognitive demand than subjective effort. It also disagrees with a large number of studies finding that objective effort often responds to demand when subjective effort does not (Downs & Crum, 1978; Hicks & Tharpe, 2002; Hornsby, 2013; Pals et al., 2013; Mackersie et al., 2015). However, a minority of studies have reported findings consistent with the current study, with subjective effort more responsive to cognitive demand than objective effort. For instance, in Feuerstein (1992), normal-hearing participants reported greater subjective effort when they experienced simulated unilateral hearing loss, but this had no effect on objective effort measured using a dual-task paradigm. In addition, Carolan et al. (2021) recently observed that although both secondary-task reaction time and effort ratings increased with greater listening demands, only subjective effort increased when financial rewards were used to increase participants' level of motivation.

The greater responsiveness of subjective effort than objective effort is unexpected theoretically. This possibility has been acknowledged before (Lemke & Besser, 2016), but no specific mechanism was proposed. Upon finding these results, Feuerstein (1992) speculated that participants may have used their task performance as a cue to rate subjective effort, while Carolan et al. (2021) simply appealed to the multidimensional nature of listening effort. If subjective ratings are being made as expected (i.e., rating the feeling of effort), it is unclear what would cause them to be elevated if not increased cognitive resource recruitment. This unexpected result could relate to the sensitivity of the measures used. As mentioned, the dual-task paradigm may have been relatively insensitive to objective effort, meaning that the genuine increase in resource recruitment may have begun at a much more favourable SNR than was found. Further, self-report measures of listening effort are generally considered the most sensitive than objective measures, exhibiting greater effect sizes (Johnson et al., 2015; Visentin et al., 2021; Giuliani et al., 2022). Importantly, such differences in sensitivity are not the same as differences in response profile. Other studies have hinted that, if functioning as intended (e.g., if sufficiently sensitive), all measures of listening effort should agree (Moore & Picou, 2018). However, if listening effort is indeed multidimensional, even very sensitive measures will differ in their responses to cognitive demand.

A perhaps more parsimonious explanation for the greater responsiveness of subjective effort is that participants were rating something else apart from the feeling of effort. Self-reports may be biased, especially since higher-level mental states are often difficult to access (Nisbett & Wilson, 1977; Althubaiti, 2016). When participants assign quantities to concepts, they may therefore rely on strategies called judgement heuristics (Tversky & Kahneman, 1974), which reduce the time and effort needed to make decisions by considering less or more-accessible information (Newell & Simon, 1972; Simon, 1990; Shah & Oppenheimer, 2008). One specific heuristic is attribute substitution, in which a difficult question is unconsciously substituted with an easier one (Kahneman & Frederick, 2002). This may occur when three conditions are met: the target attribute is difficult to assess, a related attribute is a reasonable replacement for it, and the related attribute is easier to assess than the target attribute. In the context of mental effort, Moore and Picou (2018) found evidence that participants may substitute the question "How much effort do you feel?” with “How well do you think you are performing?” In particular, a series of visual experiments demonstrated that, when task accuracy was available as a cue, it biased effort ratings, with higher accuracy cueing lower effort ratings and vice versa. This substitution meets the conditions set by Kahneman and Frederick (2002). Effort is indeed difficult to reflect on, and performance is a reasonable substitute. Further, participants can accurately assess their accuracy on speech tasks, with subjective ratings of intelligibility agreeing with behavioural measures (Cox et al., 1991; Larsby & Arlinger, 1994; Cienkowski & Speaks, 2000).

The finding that task performance biases effort ratings is consistent with a number of prior observations. For instance, self-reported effort ratings correlate better with speech intelligibility (e.g., word report scores) than with behavioural or physiological measures of listening effort (Picou et al., 2017; Fraser et al., 2010; Larsby et al., 2005; Picou & Ricketts, 2018; Zekveld et al., 2010; Downs & Crum, 1978; Feuerstein, 1992). For instance, Zekveld et al. (2010) reported significant correlations between effort ratings and intelligibility, but did not find any correlation between the effort ratings and pupil size during listening. Indeed, several of these studies (Downs & Crum, 1978; Feuerstein, 1992) concluded that effort may have been biased by performance. Further evidence comes from Zekveld and Kramer (2014), which asked participants to rate effort during speech recognition tasks with various levels of masking. The most difficult conditions had smaller pupil dilation than easier conditions, indicating that participants gave up, and indeed they rated giving up more often in these conditions. However, their effort ratings indicated that they continued to feel more effort even into the least favourable conditions. This likely reflects response bias, with lower intelligibility cueing participants to rate effort as being higher. To minimize such bias, recent studies have also asked participants to rate perceived performance, making this distinction clearer (Alhanbali et al., 2017; Carolan et al., 2021). Others have had participants rate easier-to-access feelings apart from effort, with some success. For instance, compared to effort ratings, ratings of how much listeners would like to improve the listening situation correlate better with secondary-task reaction time (Picou et al., 2017; Picou & Ricketts, 2018).

Despite the prior evidence of task performance biasing effort ratings, this is unlikely to have occurred in the current study. This is because the stimuli used were highly intelligible, with intelligibility still near ceiling even as both dimensions of listening effort increased. In addition, intelligibility was controlled for in the analyses of effort ratings. Given that intelligibility was a binary outcome, it could have driven up effort ratings for the few passages in which participants did not understand the gist. Nonetheless, this is very unlikely to fully account for the observed effect of SNR on effort rating. Although discussion of bias in effort ratings has focused on perceived performance (Moore & Picou, 2018), findings have been reported that are best explained by other forms of bias. For instance, subjective effort differs depending on the secondary task used (Wu et al., 2016) or whether participants rate effort before or after an assessment of intelligibility (Hannah et al., 2022). In reality, listeners likely use whatever cues are available to them to decide upon an effort rating (Francis & Love, 2020). I propose that another form of attribute substitution occurred in the current study, in which participants were not being biased by their perceived performance, but rather the perceived level of task difficulty. Difficulty was presumably affected by an assortment of factors, including passage-specific factors. However, the most prominent feature determining task difficulty was the SNR. The level of the noise channel in particular could have potentially been used as a cue instead, but the SNR changes were likely more noticable, since all stimuli were set to the same overall level. Thus, participants may have substituted the question of "How much effort do you feel?" with the question of "How loud does the noise seem relative to the speech?"

To be biased by task difficulty, participants may have unconsciously assumed that they were trying harder in more difficult conditions. Alternatively, they could have been influenced by demand characteristics and consciously rated effort as higher out of obligation. This substitution may have also been indirect, with greater noise causing more negative emotional responses that in turn cued greater effort (Dragovic & Giles, 2016). To achieve the effort function observed, participants may have had to accurately remember previously heard noise levels and compare them to the present passage. The just-noticeable difference for SNR is 3 dB, meaning that listeners can detect this difference 50% of the time (McShefferty et al., 2016). In contrast, the effect of SNR on effort ratings was very reliable, with each 1-dB step associated with a change in effort ratings. However, this likely only arises out of the group-level data and may not be present at the individual level. Substituting effort with difficulty would have been practical given the unavailability of performance as a cue. Even if performance did vary to a greater extent, the design of the study may have rendered it difficult to use as a cue. In particular, effort ratings were made in real time, and while noise level can be rated moment-to-moment, rating performance in this way is likely much more challenging. Finally, intelligibility itself could have been subject to bias, since this was rated subjectively. Compared to effort, this is less of a concern given that, as mentioned, participants are generally accurate in their ratings of subjective intelligibility. Nonetheless, if intelligibility ratings were biased, then it must be by a factor apart from effort or difficulty, given the unique intelligibility response profile.

4.3 Strengths, limitations, and future directions

The current study had a number of strengths when compared to prior research. The stimuli were chosen to be highly intelligible at all but the least favourable SNRs, as confirmed by the mean intelligibility scores of greater than 95% for all SNRs except for +1 dB, 0 dB, and -1 dB. This was the optimal range to find a dissociation between objective and subjective effort, since the most favourable SNRs were not yet challenging enough for both dimensions of listening effort to already be elevated. The high intelligibility also ensured that the increases in subjective effort were unlikely to be caused by participants using their performance as a cue, since intelligibility did not decrease until a much less favourable SNR was reached. To further prevent such rating bias, participants were instructed on the difference between effort, performance, and difficulty. The current study also used small manipulations of SNR, making it easy to see how objective and subjective effort (and well as intelligibility) varied in their responses to cognitive demand. In addition, this study was the first to compare objective and subjective effort measured in two separate blocks, rather than as part of the same block. This rendered the feeling of effort more accessible to participants, since they did not have to separate listening effort out from the more general effort of the dual-task paradigm.

Finally, these design strengths were enhanced by the use of a very large, high-quality sample.

Despite these many strengths, the current study also had some limitations and left several questions unaddressed. Only one measure was used for each dimension of listening effort, even though different objective and subjective measures may themselves fail to correlate well (Stand et al., 2018; Alhanbali et al., 2019). As mentioned, the measures used may have also been limited, with the secondary task potentially being too easy (Picou & Ricketts, 2014) and the effort rating scale potentially being prone to bias (Moore & Picou, 2018). In addition, participants did not rate perceived performance or difficulty, which would have further clarified the distinction between effort and these other constructs. On the objective side, future research should determine how to maximize the sensitivity of dual-task paradigms to measure objective effort, especially when intelligibility is very high. For instance, it may be necessary for studies to explicitly test, for each participant, whether all resources are being recruited even in the easiest condition. On the subjective side, studies are needed to confirm the proposed influence of perceived performance on effort ratings. This could be achieved by providing participants with inaccuracy information on the task difficulty, with the expectation that a higher stated task difficulty would be associated with higher effort ratings irrespective of performance.

For simplicity, the current study also viewed objective effort and subjective effort as distinct as well as unitary, one-dimensional constructs. However, this is likely not the case. On the objective side, the specific cognitive resources recruited during listening likely depend on the demands of the situation, such as whether background noise is present or speech is ambiguous (Rovetti et al., 2022b). On the subjective side, even "unbiased" effort ratings could be affected by factors such as the level of enjoyment derived from the materials (Herrmann & Johnsrude, 2020b). However, little research has investigated these complexities, with listening effort largely still considered a one- or two-dimensional construct. Further, the distinction between measures of objective and subjective effort may not always be clear. While behavioural and self-report measures may be relatively pure in their assessment of objective and subjective effort, respectively, peripheral physiological measures—used to assess cognitive resource recruitment—can also be influenced by emotion (Francis & Oliver, 2018). This may be especially true of the sympathetic arousal-driven pupil response, given the stronger connection of effort to parasympathetic withdrawal (Steinhauer et al., 2004; Bradley et al., 2017).

Finally, more research is needed to bring the measurement of listening effort from the lab to the clinic, contributing to the diagnosis and treatment of hearing loss. At the moment, there is no standard clinical measure of listening effort (Alhanbali et al., 2019). Although clinicians generally agree on the utility of such a measure, it is unclear how it would be used (Pichora-Fuller et al., 2016). The dimension of listening effort to be measured, for instance, may depend on the problems reported (Flake & Fried, 2020; Strand et al., 2020). If patients report cognitive deficits such as poor memory for speech, measures of objective effort may be preferred, though newer approaches may be needed that require fewer assumptions (Seeman & Sims, 2015). In contrast, if patients complain of social withdrawal or other real-world behaviours, subjective effort measures may be more suitable, since the feeling of effort may drive these behaviours (Humes, 1999; Hällgren et al., 2005; Rudner et al., 2012). Some have recommended that both dimensions of listening effort be measured to obtain a complete picture of patients' hearing health (Larsby et al., 2005; Anderson Gosselin & Gagne, 2011; Hornsby et al., 2013). Regardless of the approach, it should also be acknowledged that listening effort is not always negative per se—it is also used to achieve valued communication outcomes (Herrmann & Johnsrude, 2020a, 2020b). Indeed, a major goal of clinicians should be enabling listeners to have more of these positive listening experiences.

4.4 Conclusion and implications

In sum, the current study found different response profiles for secondary-task reaction time (i.e., objective effort), effort ratings (i.e., subjective effort), and gist understanding (i.e., speech intelligibility) as the SNR became less favourable. Subjective effort responded first, increasing even at the most favourable SNRs. Objective effort responded second, beginning to increase after an SNR of approximately +3 dB. Intelligibility responded third, beginning to decrease at an SNR less favourable than approximately +2 dB. The different effort response profiles are consistent with listening effort being multidimensional. However, the greater responsiveness of subjective effort than objective effort was unexpected from a theoretical perspective, since it is unclear what would trigger an increase in the feeling of effort if not increased cognitive resource recruitment. Indeed, these results should be interpreted with caution. The secondary task may have been too easy to be sufficiently sensitive to resource recruitment, while participants may have been biased by a cue such as perceived task difficulty in their ratings of effort. Nonetheless, the current study demonstrates that resource recruitment and the reported feeling of effort both increased as listening conditions became more challenging, and that the increase in resource recruitment may allow intelligibility to remain very high.

This study also highlights the theoretical and methodological limitations present in the listening effort literature. Future research will be required to address these limitations, as well as to determine how listening effort should be measured in a clinical setting. As mentioned, a valid and reliable measure of listening effort would be an invaluable tool for clinicians aiming to diagnose and treat otherwise-hidden cases of hearing loss. With better detection of hearing loss, cases would also be treated earlier in patients' lives, at which point they are much more likely to adapt to and benefit from hearing aids (Jolink et al., 2020). Indeed, with clinicians able to measure listening effort, they may be able to fit and adjust patients' hearing aids in such a way that minimizes effort (Wild et al., 2012). This would make hearing aids more effective for all and motivate more patients to wear them. Together, these innovations will reduce the suffering of those with hearing loss, as well as the associated economic costs. With Canada's population growing older and cases of hearing loss on the rise, this issue must be prioritized.

References

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing, 40*, 1084–1097. https://doi.org/10.1097/aud.0000 000000000697

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing, 38*(1), e39–e48. https://doi.org/10.1097/aud.0000000000000361

Althubaiti A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare, 9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Anderson Gosselin, P., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech Language and Hearing Research, 54*, 944–956. https://doi.org/10.1044/1092-4388(2010/10-0069)

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of Cognitive Hearing Science. *Scandinavian Journal of Psychology, 50*(5), 371–384. https://doi.org/10.1111/j.1467-9450.2009.00753.x

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*(2), 394–403. https://doi.org/10.1177/0013164494054002013

Ayasse, N. D., & Wingfield, A. (2018). A tipping point in listening effort: Effects of linguistic complexity and age-related hearing loss on sentence comprehension. *Trends in Hearing, 22*, 1–14. https://doi.org/10.1177/2331216518790907

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). https://doi.org/10.1016/S0079-7421(08)60452-1

Baer, T., Moore, B. C., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development, 30*, 49–72.

Bain, M. T. (2020). *Cognitive resources are recruited consistently across people during story listening* [Master's thesis, Western University]. https://ir.lib.uwo.ca/etd/7475/

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8*, 1–13. https://doi.org/10.3389/fpsyg.2017.00456

Bashford, J. A., Riener, K. R. & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & Psychophysics 51*, 211–217. https://doi.org/10.3758/BF03212247

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

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371–372. https://doi.org/10.3758/BF03328444

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: Influence of age and hearing impairment. *Brain Research Bulletin, 91*, 21–30. https://doi.org/10.1016/j.brainresbull.2012.11.005

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 1–13. https://doi.org/10.3389/fpsyg.2021.675558

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*(3), 739–753. https://doi.org/10.1111/ajps.12081

Berry, G., Hurd, L. M., Tucker, G., & Munro, D. E. (1925). Society Proceedings of the Eastern Section of the American Laryngological, Rhinological and Otological Society. *The Annals of Otology, Rhinology & Laryngology, 34*(3), 953–988. https://doi.org/10.1177/000348942503400339

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologia*, *134*(3), 330–343. https://doi.org/10.1016/j.actpsy.2010.03.006

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345. https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

Borghini, G., & Hazan, V. (2020). Effects of acoustic and semantic cues on listening effort during native and non-native speech perception. *The Journal of the Acoustical Society of America*, *147*, 3783–3794. https://doi.org/10.1121/10.0001126

Bradley, M. M., Sapigao, R. G., & Lang, P. J. (2017). Sympathetic ANS modulation of pupil diameter in emotional scene perception: Effects of hedonic content, brightness, and contrast. *Psychophysiology, 54*(10), 1419–1435. https://doi.org/10.1111/psyp.12890

Brons, I., Houben, R., & Dreschler, W. A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends in Hearing, 18*, 1–10. https://doi.org/10.1177/2331216514553924

Bridges, D., Pitiot, A., MacAskill, M. R., Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *Brain, Cognition and Mental Health*, *8*, 1–29. https://doi.org/10.7717/peerj.9414

Brown, V. A., & Strand, J. F. (2019). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition, 2*(1), 1–20. https://doi.org/10.5334/joc.89

Brown, V. A., & Strand, J. F. (2019). Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity. *Language, Cognition and Neuroscience, 34*(5), 628–640. https://doi.org/10.1080/23273798.2018.1562084

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology, 73*(9) 1431–1443. https://doi.org/10.1177/1747021820916726

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*, 1101–1109. https://doi.org/10.1121/1.1345696

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods, 50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Carolan, P. J., Heinrich, A., Munro, K. J., & Millman, R. E. (2021). Financial reward has differential effects on behavioural and self-report measures of listening effort. *International Journal of Audiology, 60*(11), 900–910. https://doi.org/10.1080/14992027.2021.1884907

Carson, R., & Beeson, C. M. L. (2013). Crossing language barriers: Using crossed random effects modelling in psycholinguistics research. *Tutorials in Quantitative Methods for Psychology, 9,* 25–41. https://doi.org/10.20982/tqmp.09.1.p025

Cienkowski, K. M., & Speaks, C. (2000). Subjective vs. objective intelligibility of sentences in listeners with hearing loss. *Journal of Speech Language and Hearing Research, 43,* 1205–1210. https://doi.org/10.1044/jslhr.4305.1205

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics, 42*(5), 476–484. https://doi.org/10.3758/BF03209755

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8*, 119s–126S. https://doi.org/10.1097/00003446-198710001-00010

Cox, R. M., Alexander, G. C., & Rivera, I. M. (1991). Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. *Journal of Speech Language and Hearing Research*, *34*(4), 904–915. doi:10.1044/jshr.3404.904

Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America*, *30*, 413–417. https://doi.org/10.1121/1.1909628

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cunningham, L. L., & Tucci, D. L. (2017). Hearing loss in adults. *New England Journal of Medicine, 377*(25), 2465–2473. https://doi.org/10.1056/NEJMra1616601

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing, 35*, 600–610. https://doi.org/10.1097/aud.0000000000000028

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports, 9*, 1–10. https://doi.org/10.1038/s41598-019-47643-1

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders*, *47*(2), 189–193. https://doi.org/10.1044/jshd.4702.189

Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech Language and Hearing Research*, *21*(4), 702–714. https://doi.org/10.1044/jshr.2104.702

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research, 42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences, 111*, 7126–7131. https://doi.org/10.1073/pnas.1318738111

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is listening in noise worth It? The neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing*, *37*, 101S–110S. https://doi.org/10.1097/aud.0000000000000300

Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expended processing capacity. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(5), 472–484. https://doi.org/10.1037/0278-7393.5.5.472

Feder, K. (2015). Prevalence of hearing loss among Canadians aged 20 to 79: Audiometric results from the 2012/2013 Canadian Health Measures Survey. *Health Reports, 26*(7), 18–25.

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *45A*(3), 353–372. https://doi.org/10.1080/02724989208250619

Feuerstein, J. F. (1992). Monaural versus binaural hearing. *Ear and Hearing, 13*(2), 80–86. https://doi.org/10.1097/00003446-199204000-00003

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, *65*, 250–259. https://doi.org/10.1007/s004260100065

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*, 456–465. https://doi.org/10.1177/2515245920952393

Fletcher, H., & Galt, R. H. (1950). The perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America, 22*, 89–151. https://doi.org/10.1121/1.1906605

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680–698. https://doi.org/10.1037/0278-7393.10.4.680

Francis, A. L., & Love, J. (2019). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), 1–27. https://doi.org/10.1002/wcs.1514

Francis, A. L., & Oliver, J. (2018). Psychophysiological measurement of affective responses during speech perception. *Hearing Research*, *369*, 103–119. https://doi.org/10.1016/j.heares.2018.07.007

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech Language and Hearing Research, 53*(1), 18–33. https://doi.org/10.1044/1092-4388(2009/08-0140)

Gagné, J., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm. *Trends in Hearing, 21*, 1–25. https://doi.org/10.1177/2331216516687287

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology, 24*(1), 63–68. https://doi.org/10.3109/03005369009077843

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology, 43*(2), 85–99. https://doi.org/10.1080/14992020400050014

Giuliani, N. P., Brown, C. J., & Wu, Y. H. (2021). Comparisons of the sensitivity and reliability of multiple measures of listening effort. *Ear and Hearing, 42*(2), 465–474. https://doi.org/10.1097/AUD.0000000000000950

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America, 22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition, 27*(2), 328–338. https://doi.org/10.3758/bf03211416

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *The American Psychologist, 59*(2), 93–104. https://doi.org/10.1037/0003-066X.59.2.93

Gurka, M. J. (2006) Selecting the best linear mixed model under REML. *The American Statistician, 60*(1), 19–26. https://doi.org/10.1198/000313006X90396

Hannah, J. A., Rovetti, J., Van Hedger, S., & Johnsrude, I. S. (2022). Task effects on self-reported listening effort [Manuscript in preparation]. Department of Psychology, Western University.

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2005). Speech understanding in quiet and noise, with and without hearing aids. *International Journal of Audiology, 44*(10), 574–583. https://doi.org/10.1080/14992020500190011

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology Human Mental Workload*, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

Hayes, M. H. S., & Patterson, D. G. (1921) Experimental development of the graphic rating method. *Psychological Bulletin, 18,* 98–99.

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*(10), 1737–1752. https://doi.org/10.1162/jocn.2008.20069

Herrmann, B., & Johnsrude, I. S. (2020a). A Model of Listening Engagement (MoLE). *Hearing Research, 397*, 1–12. https://doi.org/10.1016/j.heares.2020.108016

Herrmann, B., & Johnsrude, I. S. (2020b). Absorption and enjoyment during listening to acoustically masked stories. *Trends in Hearing, 24*, 1–18. https://doi.org/10.1177/2331216520967850

Hétu, R., Riverin, L., Lalande, N., Getty, L., & St-Cyr, C. (1988). Qualitative analysis of the handicap associated with occupational hearing loss. *British Journal of Audiology, 22*, 251–264. https://doi.org/10.3109/03005368809076462

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393–402. https://doi.org/10.1038/nrn2113

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech Language and Hearing Research, 45*(3), 573–584. https://doi.org/10.1044/1092-4388(2002/046)

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70. https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

Holmes, E., Folkeard, P., Johnsrude, I. S., & Scollie, S. (2018). Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *International Journal of Audiology*, *57*, 483–492. https://doi.org/10.1080/14992027.2018.1432901

Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends in Hearing, 20*, 1–15. https://doi.org/10.1177/2331216516667734

Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing, 34,* 523–534.

Hornsby, B. W. Y., Naylor, G., & Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear and Hearing, 37,* 136S–144S. https://doi.org/10.1097/aud.0000000000000289

Houben, R., Doorn-Bierman, M. V., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology, 52*, 753–761. https://doi.org/10.3109/14992027.2013.832415

Humes L. E. (1999). Dimensions of hearing aid outcome. *Journal of the American Academy of Audiology*, *10*(1), 26–39. https://doi.org/10.1055/s-0042-1748328

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006) Auditory measures of selective and divided attention in young and older adults using single-talker competition. *Journal of the Acoustical Society of America, 120*, 2926–2937. https://doi.org/10.1121/1.2354070

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379. https://doi.org/10.1121/1.2904826

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 240–247. https://doi.org/10.1037/0278-7393.14.2.240

Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American Journal of Audiology*, *24*, 419–431. https://doi.org/10.1044/2015_aja-14-0058

Johnsrude, I. S., & Rodd, J. M. (2016). Factors that increase processing demands when listening to speech. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 491–502). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00040-7

Jolink, C., Lansbergen, S. E., & Dreschler, W. A. (2020). Hearing disabilities and the effectiveness of rehabilitation in different age groups. *Otology & Neurotology*, *41*, e982–e988. https://doi.org/10.1097/mao.0000000000002719

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2015). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*, 1–14. https://doi.org/10.1016/j.neuron.2015.11.028

Kahneman, D. (1973). *Attention and Effort*. Prentice Hall.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker mask. *Ear and Hearing, 33*(2), 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Kramer, S. E., Kapteyn T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Audiology*, *36*(3), 155–64. https://doi.org/10.3109/00206099709071969. PMID: 9193733.

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017). Relation between listening effort and speech intelligibility in noise. *American Journal of Audiology*, *26*(3S), 378–392. https://doi.org/10.1044/2017_AJA-16-0136

Kuznetsova, A., Brockhodd, P, & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Larsby, B., & Arlinger, S. (1994). Speech recognition and just-follow-conversation tasks for normal-hearing and hearing-impaired listeners with different maskers. *International Journal of Audiology, 33*(3), 165–176. https://doi.org/10.3109/00206099409071877

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S, (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology, 44*(3), 131–143. https://doi.org/10.1080/14992020500057244

Lau, M. K., Hicks, C., Kroll, T., & Zupancic, S. (2019). Effect of auditory task type on physiological and subjective measures of listening effort in individuals with normal hearing. *Journal of Speech, Language, and Hearing Research, 62*, 1549–1560. https://doi.org/10.1044/2018_jslhr-h-17-0473

Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Research, 36*(3), 291–298. https://doi.org/10.1016/0165-1781(91)90027-m

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort. *Ear and Hearing*, *37*, 77S–84S. https://doi.org/10.1097/aud.0000000000000304

Lin, F. R., Yaffe, K., Xia, J., Xue, Q., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., & Simonsick, E. M. (2013). Hearing loss and cognitive decline among older adults. *JAMA Internal Medicine, 173*, 1–14. https://doi.org/10.1001/jamainternmed.2013.1868

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Luke, S. G. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., & Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *Journal of the Acoustical Society of America, 127*, 1491–1505. https://doi.org/10.1121/1.3299168

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology, 22*, 113–122. https://doi.org/10.3766/jaaa.22.2.6

Mackersie, C. L., Macphee, I. X., & Heldt, E. W. (2015). Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear and Hearing, 36*, 145–154. https://doi.org/10.1097/aud.0000000000000091

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1–23. https://doi.org/10.3758/s13428-011-0124-6

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., Meyer-Lindenberg, A., Goldberg, T. E., Callicott, J. H., & Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters*, *392*, 32–37. https://doi.org/10.1016/j.neulet.2005.09.025

Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear and Hearing, 37*, 28S–34S. https://doi.org/10.1097/aud.0000000000000292

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*, 953–978. https://doi.org/10.1080/01690965.2012.705006

McAuliffe, M. J., Wilding, P. J., Rickard, N. A., & O'Beirne, G. A. (2012). Effect of speaker age on speech recognition and perceived listening effort in older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 55*, 838–847. https://doi.org/10.1044/1092-4388(2011/11-0101)

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A, 58*, 22–33. https://doi.org/10.1080/02724980443000151

McDaid, D., Park, A., & Chadha, S. (2021). Estimating the global costs of hearing loss. *International Journal of Audiology, 60*, 162–170. https://doi.org/10.1080/14992027.2021.1883197

McGarrigle, R., Rakusen, L., & Mattys, S. (2020). Effortful listening under the microscope: Examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiology, 0*, 1–22. https://doi.org/10.1111/psyp.13703

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology, 53*, 433–445. https://doi.org/10.3109/14992027.2014.890296

McLaughlin. D. J., Braver, T. S., Peelle, J. E. (2021). Measuring the subjective cost of listening effort using a discounting task. *Journal of Speech Language and Hearing Research, 64*(2), 337–347. https://doi.org/10.1044/2020_JSLHR-20-00086

McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *The Quarterly Journal of Experimental Psychology, 29*(4), 651–667. https://doi.org/10.1080/14640747708400639

McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2016). The just-meaningful difference in speech-to-noise ratio. *Trends in Hearing, 20*, 1–11. https://doi.org/10.1177/2331216515626570

Michalek, A. M. P., Ash, I., & Schwartz, K. (2018). The independence of working memory capacity and audiovisual cues when listening in noise. *Scandinavian Journal of Psychology, 59*(6), 578–585. https://doi.org/10.1111/sjop.12480

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *Journal of Speech, Language, and Hearing Research, 61*(9), 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451

Morys-Carter, W.L. (2021, April 26). *Participant IDs for Pavlovia*. VESPR. https://moryscarter.com/vespr/pavlovia.php

Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Nicholson, N. R. (2009). Social isolation in older adults: An evolutionary concept analysis. *Journal of Advanced Nursing, 65*(6), 1342–1352. https://doi.org/10.1111/j.1365-2648.2008.04959.x

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C., Schwab, & H. C., Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. I. Speech perception* (pp. 113–158). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-631403-8.50009-6

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage, 55*, 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common α oscillatory network. *The Journal of Neuroscience, 32*(36), 12376–12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., & Kramer, S. E. (2017a). Effects of hearing impairment and hearing aid amplification on listening effort. *Ear and Hearing, 38*, 267–281. https://doi.org/10.1097/aud.0000000000000396

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., N. J. Versfeld, & Kramer, S. E. (2017b). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research, 351*, 68–79. https://doi.org/10.1016/j.heares.2017.05.012

Panico, J., & Healey, E. C. (2009). Influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. *Journal of Speech, Language, and Hearing Research*, *52*(2), 534–546. https://doi.org/10.1044/1092-4388(2008/07-0238)

Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech Language and Hearing Research, 56*(4), 1075–1084. https://doi.org/10.1044/1092-4388(2012/12-0074)

Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., & Polley, D. B. (2020). Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *ELife, 9*, e51419. https://doi.org/10.7554/eLife.51419

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peelle, J. E. (2018). Listening Effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*, 204–214. https://doi.org/10.1097/aud.0000000000000494

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-021-01694-3

Picou, E. M., Bean, B., Marcrum, S. C., Ricketts, T. A., & Hornsby, B. W. Y. (2019). Moderate reverberation does not increase subjective Fatigue, subjective listening effort, or behavioral listening effort in school-aged children. *Frontiers in Psychology, 10*, 1–16. https://doi.org/10.3389/fpsyg.2019.01749

Picou, E. M., & Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear and Hearing, 35*(6), 611–622. https://doi.org/10.1097/AUD.0000000000000055

Picou, E. M., & Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *International Journal of Audiology, 57*(6), 457–467. https://doi.org/10.1080/14992027.2018.1431696

Picou, E. M., Moore, T. M., & Ricketts, T. A. (2017). The effects of directional processing on objective and subjective listening effort. *Journal of Speech Language and Hearing Research, 60*(1), 199–211. https://doi.org/10.1044/2016_JSLHR-H-15-0416

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2013). How hearing aids, background Noise, and visual cues influence objective listening effort. *Ear and Hearing, 34*(5), e52–e64. https://doi.org/10.1097/AUD.0b013e31827f0431

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research, 54*(5), 1416–1430. https://doi.org/10.1044/1092-4388(2011/10-0154)

Pienkowski, M. (2017). On the etiology of listening difficulties in noise despite clinically normal audiograms. *Ear and Hearing, 38*(2), 135–148. https://doi.org/10.1097/AUD.0000000000000388

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology, 42*, 26–32. https://doi.org/10.3109/14992020309074641

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K.L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing, 37*, 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, cognition, and healthy aging: Social and public health implications of the links between age-related declines in hearing and cognition. *Seminars in Hearing, 36*, 122–139. https://doi.org/10.1055/s-0035-1555116

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual consequences of "hidden" hearing loss. *Trends in Hearing, 18*, 1–11. https://doi.org/10.1177/2331216514550621

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America, 57*, S5. https://doi.org/10.1121/1.1995329

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* (Report No. DAMTP 2009/NA06). https://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 71*(2), 264–272. https://doi.org/10.1037/h0022853

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20*(3), 241–248. https://doi.org/10.1080/14640746808400158

Rabbitt, P. M. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica, 111*, 167–176. https://doi.org/10.3109/00016489109127274

Reuter-Lorenz, P. A., Stanczak, L., & Miller, A. C. (1999). Neural recruitment and cognitive aging: Two hemispheres are better than one, especially as you age. *Psychological Science, 10*, 494–500. https://doi.org/10.1111/1467-9280.00195

Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology, 11*, 374–385. https://doi.org/10.1017/S1138741600004406

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America, 136*(5), 2642–2653. https://doi.org/10.1121/1.4897398

Richter, M., Gendolla, G.H.E., & Wright, R.A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 149–186). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2016.02.001

Ritz, H., Wild, C. J., & Johnsrude, I. J. (2022). Parametric cognitive load reveals hidden costs in the neural processing of perfectly intelligible degraded speech. *Journal of Neuroscience, 42*, 4619–4628. https://doi.org/10.1523/JNEUROSCI.1777-21.2022

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*(8), 1261–1269. https://doi.org/10.1093/cercor/bhi009

Rodd. J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain & Language, 115*(3), 182–188. https://doi.org/10.1016/j.bandl.2010.07.005

Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional near-infrared spectroscopy as a measure of listening effort in older adults who use hearing aids. *Trends in Hearing*, 23, 1–22. https://doi.org/10.1177/2331216519886722

Rovetti, J., Goy, H., & Russo, F.A. (2022a). Reduced semantic context and signal-to-noise ratio increase listening effort as measured using functional near-infrared spectroscopy. *Ear and Hearing, 43*, 836–848. https://doi.org/10.1097/AUD.0000000000001137

Rovetti, J., Davis, M. H., Rodd, J. M., Hakyemez, H., Lee, S. M., Chau, A., Johnsrude, I. S. (2022b). *The brain responds differently to acoustic and linguistic challenges to speech comprehension* [Manuscript in preparation]. Department of Psychology, Western University.

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology, 23*, 577–589. https://doi.org/10.3766/jaaa.23.7.7

Rudner, M., Lyberg-Åhlander, V., Brännström, J., Nirme, J., Pichora-Fuller, M. K., & Sahlén, B. (2018). Listening comprehension and listening effort in the primary school classroom. *Frontiers in Psychology, 9*, 1–7. https://doi.org/10.3389/fpsyg.2018.01193

Rudner, M., Mishra, S., Stenfelt, S., Lunner, T., & Rönnberg, J. (2016). Seeing the talker's face improves free recall of speech for young adults with normal hearing but not older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 59*(3), 590–599. https://doi.org/10.1044/2015_JSLHR-H-15-0014

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech Language and Hearing Research, 52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Saleh, H. K., Folkeard, P., Macpherson, & E., Scollie, S. (2020). Adaptation of the Connected Speech Test: Rerecording and passage equivalency. *American Journal of Audiology, 29*(2), 259–264. https://doi.org/10.1044/2019_AJA-19-00052

Satopää, V., Albrecht, J., Irwin, D., & Raghava, B. (2011). Finding a “kneedle” in a haystack: Detecting knee points in system behavior. *Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171. https://doi.org/10.1109/ICDCSW.2011.20

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114. https://doi.org/10.2307/3002019

Schepker, H., Haeder, K., Rennies, J., & Holube, I. (2016). Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners. *International Journal of Audiology, 55*(12), 738–747. https://doi.org/10.1080/14992027.2016.1219774

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage, 44*(3), 1133–1143. https://doi.org/10.1016/j.neuroimage.2008.09.045

Shtyrov, Y. (2010). Automaticity and attentional control in spoken language processing: Neurophysiological evidence. *The Mental Lexicon, 5*, 255–276. https://doi.org/10.1075/ML.5.2.06SHT

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19. https://doi.org/10.1146/annurev.ps.41.020190.000245

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology, 26*(2), 183–196. https://doi.org/10.3766/jaaa.26.2.7

Sommers, M. S., & Phelps, D. (2016). Listening effort in younger and older adults. *Ear and Hearing, 37*, 62S–68S. https://doi.org/10.1097/AUD.0000000000000322

Staal, M.A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework* (Report No. 2004-212824). https://ntrs.nasa.gov/

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology, 52*(1), 77–86. https://doi.org/10.1016/j.ijpsycho.2003.12.005

Strand, J., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech Language and Hearing Research*, *61*, 1–24. https://doi.org/10.1044/2018_jslhr-h-17-0257

Strand, J. F., Ray, L., Dillman-Hasso, N. H., Villanueva, J., & Brown, V. A. (2020). Understanding speech amid the jingle and jangle: Recommendations for improving measurement practices in listening effort research. *Auditory Perception & Cognition, 3*, 169-188. https://doi.org/10.1080/25742442.2021.1903293

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 809–825. https://doi.org/10.3758/s13415-017-0513-0

Sucala, M., Scheckner, B., & David, D. (2010). Psychological time: Interval length judgments and subjective passage of time judgments. *Current Psychology Letters: Behaviour, Brain and Cognition, 26*(2), 1–9. https://doi.org/10.4000/cpl.4998

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77,* 184–197. https://doi.org/10.1016/j.chb.2017.08.038

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. The Science Press. https://doi.org/10.1037/13283-000

Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., & Cruickshanks, K. J. (2015). Self-reported hearing difficulties among adults with normal audiograms: The Beaver Dam Offspring Study. *Ear and Hearing, 36*(6), e290–e299. https://doi.org/10.1097/AUD.0000000000000195

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761–766. https://doi.org/10.1037/a0014802

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Van Hedger, S., & Johnsrude, I. S. (2021). Speech perception under adverse listening conditions. In L. L. Holt, J. E. Peelle, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Speech perception* (pp. 141–171). ASA Press. https://doi.org/10.1007/978-3-030-81542-4_6

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* [Doctoral dissertation, Eindhoven University of Technology]. https://pure.tue.nl/ws/files/3389175/152538.pdf

Visentin, C., Valzolgher, C., Pellegatti, M., Potente, P., Pavani, F., & Prodi, N. (2022). A comparison of simultaneously-obtained measures of listening effort: Pupil dilation, verbal response time and self-rating. *International Journal of Audiology, 61*(7), 561–573. https://doi.org/10.1080/14992027.2021.1921290

Walker, J. J. (2013). Audiometry screening and interpretation. *American Family Physician, 87*(1), 41–47. https://www.aafp.org/pubs/afp/issues/2013/0101/p41.html

Watson C. S., Kelly W. J., & Wroton, H. W. (1976) Factors in the discrimination of tonal patterns II: Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America, 60*, 1176–1186. https://doi.org/10.1121/1.381220

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology, 7*, 1–12. https://doi.org/10.3389/fpsyg.2016.00345

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research, 369*, 67–78. https://doi.org/10.1016/j.heares.2018.05.006.

Westbrook, A., Kester, D., Braver, T. S., & Pessiglione, M. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLOS ONE, 8*(7), 1–8. https://doi.org/10.1371/journal.pone.0068210

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*(3), 475–482. https://doi.org/10.1007/BF03395630

Whitehead, A. N. (1925). *Science and the Modern World*. The Macmillan Company.

Wickens, C. D. (2008). Cognitive resources and mental workload. *Human Factors, 50*, 449–455 https://doi.org/10.1518/001872008X288394.

Wijayasiri, P., Hartley, D. E., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research, 351*, 55–67. https://doi.org/10.1016/j.heares.2017.05.010

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience, 32*, 14010–14021. https://doi.org/10.1523/jneurosci.1528-12.2012

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends in Hearing, 25*, 1–26. https://doi.org/10.1177/23312165211027688

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1–32. https://doi.org/10.1177/2331216518800869

Wright, R. A. (2014). Presidential address 2013: Fatigue influence on effort—Considering implications for self-regulatory restraint. *Motivation and Emotion, 38*, 183–195. https://doi.org/10.1007/s11031-014-9406-5

Wu, Y. H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., & Bentler, R. (2014). Measuring listening effort: Driving simulator versus simple dual-task paradigm. *Ear and Hearing*, *35*(6), 623–632. https://doi.org/10.1097/AUD.0000000000000079

Wu, Y., Stangl, E., Zhang, X., Perkins, J., Eilers, E. (2016). Psychometric function of dual-task paradigms for measuring listening effort. *Ear and Hearing, 37*(6), 660–670. https://doi.org/10.1097/AUD.0000000000000335

Yeh, Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30*(1), 111–120. https://doi.org/10.1177/001872088803000110

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology, 51*(3), 277–284. https://doi.org/10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing, 32*(4), 498–510. https://doi.org/10.1097/AUD.0b013e31820512bb

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*(4), 480–490. https://doi.org/10.1097/aud.0b013e3181d4f251

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013) The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *Journal of the Acoustical Society of America, 134*, 2225–2234. https://doi.org/10.1121/1.4817926

Appendices

Appendix A: Connected Speech Test passage sentences

GRASS

Grass **CAN GROW** in all climates.
THERE are many forms of grasses.
MANY GRASSES are important food **SOURCES**.
Some grasses **GROW** higher than a **MAN'S HEAD**.
AMONG THESE are bamboo and sugar cane.
Other types are **ONLY** a **FEW INCHES TALL**.
Some grasses **ARE AS SLENDER** as threads.
Others are stiff enough to **STAND** a heavy **SNOW**.
MOST grasses are **FLOWERING PLANTS**.
These flowers bloom **MAINLY** in the **SPRING**

WOLF

The wolf **IS** a **MEMBER** of the **DOG** family.
A wolf **LOOKS** like a **SKINNY** wild **DOG**.
It has a **WIDE HEAD** and pointed **NOSE**.
Wolves **LIVE** in North **AMERICA**, Europe, and Asia.
Wolves **USED** to **LIVE** all over the United **STATES**.
GRAY wolves are **SELDOM SEEN** nowadays.
THEY live in the Rockies and **NORTHERN** states.
Wolves **HUNT** in packs and **MATE** for **LIFE**.
The average wolf pack consists of **TEN** wolves.
A female wolf gives **BIRTH** every other **YEAR**.

EYE

The eye is a most **IMPORTANT SENSE** organ.
We **USE** it to **VIEW** the **WORLD**.
ALMOST EVERY ACTIVITY INVOLVES the eyes.
EYES are **OUR** windows to the **WORLD**.
The **LENS** of the eye collects **LIGHT**.
The **LIGHT** is **FOCUSED INSIDE** the eye.
This information is sent to the **BRAIN**.
The brain then begins to **PROCESS** the **IMAGE**.
Eyes help us to enjoy **BOOKS** and **PAINTINGS**.
We **SEE** beauty in **MOUNTAINS** and **SUNSETS**.

WHEAT

Wheat is a **CHIEF SOURCE** of food.
MILLIONS of **PEOPLE DEPEND** on wheat **PRODUCTS**.
It is the most **WIDELY** used human **FOOD**.
Americans **PRIZE** wheat **MORE HIGHLY** than **OTHER** grains.
Wheat is **GROWN** on the **PLAINS** of the United States.
More wheat is **PRODUCED** there than **RICE**.
However, rice is **CHEAPER** to **PRODUCE**.
It **CAN** be **PLANTED** and **HARVESTED** by **HAND**.
Rice is **IMPORTANT** to **OVERPOPULATED** countries.
It is their **PRIMARY** source of nutrition.

CLOCK

Clocks are **INSTRUMENTS** that can **MEASURE** time.
They **DIVIDE** days into regular **INTERVALS**.
Originally, **TREE SHADOWS** were **USED** to mark time.
The **SHORTEST** shadows **OCCUR** around midday.
LONGER shadows occur in morning and **LATE AFTERNOON**.
The **FIRST** clock invented was the **SUNDIAL**.
LATER, the water clock was **DEVELOPED** in **CHINA**.
It could **MEASURE** time on **CLOUDY** days.
WATER clocks were used for several **THOUSAND YEARS**.
EARLY GREEKS and Romans **ALSO** used clocks.

WEED

Weeds are considered **WORTHLESS PLANTS**.
The **DIFFERENCE BETWEEN** weeds and useful **PLANTS** is unclear.
WHERE a **WEED GROWS** determines its usefulness.
OATS GROWING in a **CORNFIELD** are considered weeds.
Oats growing in an **OATFIELD** are useful **PLANTS**.
Much crop damage is **CAUSED BY** weeds.
Experts estimate it at **FIVE** dollars per person.
FARMERS SPEND THOUSANDS of **DOLLARS** for **WEED** sprays.
Chemicals used to **KILL** weeds can be harmful.

EGG

Many kinds of animals and **BIRDS PRODUCE** eggs.
The **MAIN PURPOSE** of eggs is to breed **YOUNG**.
Most young **ANIMALS BEGIN** as an **EGG**.
PEOPLE usually think of the egg as a **FOOD**.
Actually, **FEW** kinds of eggs are **EATEN**.
Bird's eggs are **LARGER** than **MAMMAL'S**.
Their eggs **CONTAIN FOOD** for the young **BIRD**.
Young birds **DEVELOP OUTSIDE** the mother's **BODY**.
The ostrich **EGG** is the **LARGEST** type.
The **HUMAN EGG** is **ONE** of the smallest.

EAR

The ear is an important **SENSE ORGAN**.
The ear **HAS** two main **PURPOSES**.
It lets **MAN HEAR** and **MAINTAIN** his balance.
GOOD hearing permits **PEOPLE** to understand **SPEECH**.
Through speech, we **EXCHANGE** ideas and **OPINIONS**.
HEARING ALSO makes man **AWARE** of **DANGER**.
The ear's **BALANCE** mechanism helps us walk **UPRIGHT**.
DAMAGE to this section causes **STAGGERING**.
The **PERSON** also **GETS** disoriented and **DIZZY**.
This kind of dizziness is **CALLED VERTIGO**.

KITE

A kite is **FLOWN** at the **END** of a string.
It is made of paper on a **LIGHT FRAME**.
Kites **MAY** be **SHAPED** like **DRAGONS** or birds.
The **KITE** was **INVENTED** two thousand years **AGO**.
HISTORIANS THINK the kite was invented in **GREECE**.
The Chinese **CLAIM** that they **INVENTED** the kite.
They **ARGUE** that it was used in **WARS**.
In **CHINA** a day is set **ASIDE** as **KITE'S** day.
Kite's day **FALLS** on the **SEVENTH** of **JULY**.

LAKE

Lakes are **BODIES** of water **SURROUNDED** by **LAND**.
They are **LOCATED** in **EVERY** large **COUNTRY**.
Some lakes are **FOUND** at **EXTREMELY** high altitudes.
Others are many feet **BELOW SEA** level.
MOUNTAIN lakes were **FORMED** by glaciers.
The **WORD** lake **MEANS** a large pond or **HOLE**.
The Caspian **SEA** is **REALLY** a lake.
SOME other well **KNOWN** seas are also lakes.
Lakes affect the weather for **MANY MILES AROUND**.
PEOPLE USE lakes for **RECREATION** and industry.

GOLD

Gold was one of the first known **METALS**.
For **MANY YEARS** gold has **SYMBOLIZED WEALTH**.
EVEN the early cave man knew **ABOUT** gold.
ANCIENT EGYPTIANS hammered gold into **LEAVES**.
They used these leaves to **DECORATE** their **TOMBS**.
A **SCIENCE** grew up around efforts to make gold.
It **STARTED DURING** the **MIDDLE** ages.
The ancient scientists **NEVER ACHIEVED** their **GOAL**.
Modern **SCIENTISTS** have made these **DREAMS** come **TRUE**.
THEY now **MAKE** gold by a **CHEMICAL** process.

OYSTER

Oysters are animals that live in **SEA** shells.
The oyster lives in many **PARTS** of the **WORLD**.
It **LIVES MOSTLY** in quiet, shallow **WATERS**.
It **IS MAN'S** most **VALUABLE SEAFOOD**.
The oyster's **SHELL** forms a **SHELTER**.
The **SHELL** is divided into two halves.
They are **FASTENED TOGETHER** at **ONE END**.
The left **HALF** is larger and **THICKER**.

<u>ZIPPER</u>

A zipper is any kind of **SLIDE FASTENER**.
ALL zippers **HAVE** two rows of **TEETH**.
The two **EDGES** of the zipper fasten **TOGETHER**.
The **TEETH HOLD** the zipper **TOGETHER**.
The edges **STAY** fastened **TILL** they are **RELEASED**.
They are released **BY DRAWING** the slide back.
Slide zippers are **OFTEN** used to **FASTEN CLOTHING**.
They **ARE USED** on **LUGGAGE** and briefcases.
The **FIRST** zipper was invented by an **AMERICAN**.
It **WAS** made of connected **HOOKS** and eyes.

<u>LEMON</u>

A **LEMON** is an oval, yellow citrus **FRUIT**.
It **GROWS** in Southern California and **FLORIDA**.
Lemon trees are **MEDIUM** sized, **WITH SPREADING BRANCHES**.
They have **PALE GREEN LEAVES** and large flowers.
The flowers are **WHITE**, with **PURPLE UNDERNEATH**.
The lemon **FLOWER** smells sweet.
SOME types of lemons have **NO** seeds.
OTHER types have **MANY** seeds.
Their **FRUIT** is a **SPECIAL TYPE** of **CITRUS**.
It usually has a **SOUR TASTE**.

<u>GLOVE</u>

Gloves are **CLOTHING WORN ON** the **HANDS**.
The **WORD "GLOVE" MEANS** "palm of the hand".
CRUDE GLOVES were **WORN** by **PRIMITIVE MAN**.
Greeks wore **WORKING** gloves to **PROTECT** their hands.
The **ROMANS USED** gloves as a sign of **RANK**.
Knights used to fasten gloves to their helmets.
The gloves **SHOWED** their **DEVOTION** to their **LADIES**.
A glove thrown on the **GROUND SIGNALED** a challenge.

DONKEY

Donkeys are **SMALLER**, sturdier relatives of **HORSES**.
The **WILD** donkey is **SHAPED** like a **ZEBRA**.
It is four **FEET** high at the **SHOULDERS**.
The donkey's **COAT** is **GRAY** and black.
It **HAS** a **DARK LINE** along its **BACK**.
This **ANIMAL** is **EXTREMELY INTELLIGENT**.
SURPRISINGLY, it is also a **SWIFT RUNNER**.
Man has **TAMED** donkeys for his personal use.
Donkeys are **OFTEN** used as **BEASTS** of burden.
All donkeys are **NOTED** for their **HUGE EARS**

EAGLE

The eagle is a large bird of **PREY**.
It has powerful **WINGS** and **SHARP** eyes.
The **EAGLE** is a **SYMBOL** of courage and freedom.
The **BALD** eagle is America's **NATIONAL BIRD**.
THERE are **SEVERAL** different kinds of eagles.
Each **TYPE IS** very **DIFFERENT** in **SIZE** and color.
Eagles **HAVE** strong beaks and **POWERFUL CLAWS**.
The eagle's **BEAK** is as long as its **HEAD**.
The beak's upper **HALF** hooks over the **LOWER**.
The eagle **USES** its **POWERFUL** beak to **CATCH** its **PREY**.

VIOLIN

The violin is the best **KNOWN** stringed **INSTRUMENT**.
EARLY VIOLINS did not produce clear tones.
These violins were **VERY ROUGH SOUNDING**.
LATER violin **MAKERS** improved their craft.
Their **VIOLINS** were **EXTREMELY** well made.
The **VIOLIN BECAME** an **INSTRUMENT** for beautiful **MUSIC**.
Only **SMALL CHANGES** have occurred in violin **DESIGN**.
Violins must be **MADE** with **GREAT** care.
The **WOOD USED** greatly influences the tone.

WOODPECKER

The woodpecker is a bird with a **STRONG BEAK**.
It bores **HOLES** in **TREES** looking for **INSECTS**.
Woodpeckers **LIVE** in all parts of the world.
The **TOES** of woodpeckers **ARE VERY UNUSUAL**.
Two **POINT FORWARD** and two face **BACKWARD**.
This allows the **BIRD** to cling to **TREES**.
The **TAIL FEATHERS** of a woodpecker are **STIFF**.
THEY can **USE** their tails as a **SUPPORT**.
They also use their tails to grasp **TREES**.
Woodpeckers **HAVE** long **TONGUES** with pointed **TIPS**.

LAWN

A lawn is an **AREA** planted **WITH** grass.
GREEN, trimmed lawns are a beautiful **SIGHT**.
People **LIKE** to plant lawns around their **HOMES**.
Hospitals **OFTEN HAVE** lawns **AROUND** them.
MOST public **BUILDINGS** have **LAWNS**.
Lawns **HELP** to keep **SOIL** from eroding.
A **GOOD** lawn is **VERY** thickly **PLANTED**.
There are **FOUR** hundred plants **PER** square **FOOT**.
EACH plant has several **BLADES** of grass.
There are several **DIFFERENT KINDS** of **GRASSES**.

LEAD

Lead is a **SOFT**, **HEAVY**, metallic element.
It is **OFTEN** combined with other **METALS**.
MANY USEFUL OBJECTS contain some lead **MIXTURE**.
The Romans **USED LEAD** for **WATER PIPES**.
Their **PUBLIC** baths were lined **WITH** lead.
The **WORD** "plumber" means a **WORKER** in lead.
Lead is **ONE** of the **HEAVIEST KNOWN** metals.
It is **ELEVEN** times as **HEAVY** as **WATER**.
The **EXPRESSION** "as **HEAVY** as lead" is common.

WINDOW

Windows **PROVIDE LIGHT** and air to **ROOMS**.
Windows were **ONCE COVERED** with **CRUDE SHUTTERS**.
Later, oiled **PAPER** was **USED** for windowpanes.
GLASS windows **FIRST** appeared in ancient Rome.
COLORED glass was used in European **WINDOWS**.
SOME CHURCHES were **FAMOUS** for their **BEAUTIFUL** windows.
These windows **DISPLAYED PICTURES** from the **BIBLE**.
PIECES of glass were **HELD** together by lead.
SUCH windows **MAY** be seen in French cathedrals.
English churches also contain **STAINED** glass windows.

CHIMNEY

A chimney **CARRIES SMOKE** from a **FIREPLACE**.
It **ALSO SUPPLIES** the fire with **OXYGEN**.
Warm air is **LIGHTER** than **COLD** air.
Warm air **ABOVE** the fire **TENDS** to rise.
As the **WARM** air **RISES**, cold air rushes in.
A draft is **CREATED** in the **CHIMNEY**.
The draft **PROVIDES** the oxygen **NEEDED** for the **FIRE**.
Chimneys must **STAND HIGHER** than the **BUILDING**.
Otherwise, the chimney **WILL** not **DRAW PROPERLY**.
CHIMNEYS can **IMPROVE** the appearance of a home.

LIME

Limes are **CLOSELY** related to the lemon.
They **TASTE** more **SOUR THAN** the lemon.
Lime **SKIN** is **THICKER** than the **LEMON'S**.
Limes **GROW** on **SMALL CITRUS** trees.
These trees **GROW** in **MORE** tropical **AREAS**.
The **MAJORITY** of limes are **PRODUCED** in Florida.
Lime trees **GROW** to be **TEN** feet **TALL**.
Limes are **PRINCIPALLY** used for making **JUICE**.

KANGAROO

The kangaroo **CARRIES** its **YOUNG** in a **POUCH**.
The pouch is **LOCATED** outside of the **ABDOMEN**.
ANIMALS with **POUCHES** are not found in **AMERICA**.
The kangaroo's **NATIVE COUNTRY** is **AUSTRALIA**.
There are many different kinds of **KANGAROOS**.
The **SMALLEST ARE** the same size as a **RABBIT**.
The largest **ARE NEARLY** seven feet tall.
Their back **LEGS** are larger than their **FRONT** legs.
Kangaroo fossils have **RECENTLY BEEN FOUND**.
Prehistoric kangaroos **GREW** to **BE** very **LARGE**.

OWL

Owls **HUNT** alone at **NIGHT** for food.
THESE BIRDS kill and **EAT** small **ANIMALS**.
They are **BIRDS** of prey, like **EAGLES**.
OWLS defend our **GARDENS** by eating **MICE**.
They are **CLOSELY** related to night **HAWKS**.
There are five **HUNDRED** different **KINDS** of owls.
They live throughout **COLD** and **TROPICAL** climates.
Owls **USUALLY** live **ALONE** in the **FOREST**.
SOMETIMES they exist on remote **SEA** islands.
Owls are **KNOWN FOR** their **SOLEMN** expression.

KNIFE

The knife is a very **HELPFUL UTENSIL**.
It was a **TOOL** developed by the **CAVE** man.
He **SHARPENED** pieces of **STONE** to **MAKE** knives.
The knives were **USED** for **SKINNING** and **CUTTING** meat.
The knife **ALSO SERVES** as a **WEAPON**.
MANY different knives are used in **INDUSTRY** today.
DIFFERENT knives are used for **VARIOUS TASKS**.
The knife has **BEEN** used to **CREATE ARTWORK**.
ARTISTS use **KNIVES** to paint pictures.

CABBAGE

Cabbage is the **MOST COMMON** garden **VEGETABLE**.
It has **THICK LEAVES** which curl inward.
They form a **ROUND HEAD** eight inches **ACROSS**.
The **WORD** cabbage is Latin for "**HEAD**".
The **CABBAGE** plant can live through **SEVERAL FREEZES**.
It also **GROWS** in the heat of **SUMMER**.
EARLY SPRING cabbage is **PLANTED** in greenhouses.
This protects the **YOUNG PLANTS FROM FROST**.
AFTER six **WEEKS** they may be moved outdoors.
TRANSPLANTING is done before the end of spring.

Appendix B: Passage comprehension questions

Passage Topic	Target	Foil
Grass	Bamboo and sugar cane are types of grasses.	Grass is one of the largest plant families.
Wolf	Wolves can be found across multiple continents.	Wolves can track their prey for hours at a time.
Eye	The eye's lens is responsible for collecting light.	Eyes are protected from debris by the eyelashes.
Wheat	The United States produces more wheat than rice.	China is the world's top wheat producer.
Clock	The sundial was the first type of clock.	Modern digital clocks run on electricity.
Weed	Weeds lead to high costs related to crop damage.	Services can be hired to help kill weeds.
Egg	Birds produce bigger eggs than mammals.	Chicken eggs are most commonly eaten by humans.
Ear	The ear has three bones.	The ear is involved in balance.
Kite	China argues that they invented the kite.	Kites have been used as a military signalling device.
Lake	Canada is the country with the most lakes.	Some lakes are well below sea level.
Gold	Ancient Egyptians used gold to decorate tombs.	Gold is an excellent conductor of heat and electricity.
Oyster	Oyster shells have two halves.	Some oysters produce pearls.
Zipper	Slide zippers can be used for clothing.	Zippers are often made of plastic and metal.
Lemon	Lemon flowers are white and purple.	Lemons are native to Asia.
Glove	Knights often attached gloves to their helmets.	Gloves can be used to keep the hands warm.
Donkey	Donkeys are very good at running.	Donkeys can breed with horses to produce mules.
Eagle	Eagles mate with each other for life.	Types of eagles vary considerably in size and colour.
Violin	Violin design has not changed very much in recent years.	The first violins were made in Italy.
Woodpecker	Woodpeckers' toes face different directions.	Woodpeckers have fuzzy noses.
Lawn	Lawns help to prevent soil erosion.	Lawns around buildings are regularly cut.
Lead	Roman public baths were lined with lead.	Poisoning from lead exposure can be fatal.
Window	Ancient Rome was first to use glass windows.	Modern windows are more resistant to shattering.
Chimney	Warm air from fires rises up through chimneys.	Most chimneys have a protective inner layer.
Lime	Limes sink in water while lemons float.	Lime trees mostly grow in tropical areas.
Kangaroo	Some kangaroos are as small as rabbits.	A group of kangaroos is called a mob.
Owl	Owls are a relative of the night hawk.	Owls' wings are designed for silent flight.
Knife	Knives can be useful for working with meat.	Knives can be used by doctors for surgery.
Cabbage	Cabbage is resistant to cold temperatures.	Cabbage is used to make sauerkraut.

Appendix C: Ethics approval

Date: 17 March 2022
To: Dr. Ingrid Johnsrude
Project ID: 119968
Study Title: Behavioural Studies of Listening Effort and Intelligibility
Short Title: Listening Effort and Intelligibility
Application Type: NMREB Initial Application
Review Type: Delegated
Full Board Reporting Date: 01/Apr/2022
Date Approval Issued: 17/Mar/2022 16:07
REB Approval Expiry Date: 17/Mar/2023

Dear Dr. Ingrid Johnsrude

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. **All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study.**

Documents Approved:

Document Name	Document Type	Document Date	Document Version
2_SIN_subjective_and_objective_effort_tasks	Other Data Collection Instruments	29/Nov/2021	1
demographics	Online Survey	29/Nov/2021	1
1-Audio_calibration	Other Data Collection Instruments	29/Nov/2021	1
3-SiN_audiovisual	Other Data Collection Instruments	29/Nov/2021	1
4-SiN_intelligbility	Other Data Collection Instruments	29/Nov/2021	1
5-progressive_matrices	Other Data Collection Instruments	29/Nov/2021	1
6-2-back_task	Other Data Collection Instruments	29/Nov/2021	1
7-Mill_Hill_vocab	Other Data Collection Instruments	29/Nov/2021	1
8-mental_rotation	Other Data Collection Instruments	29/Nov/2021	1
9-2-back_auditory	Other Data Collection Instruments	30/Nov/2021	1
10-sentence_repetition	Other Data Collection Instruments	30/Nov/2021	1
SONA_Recruitment_C	Recruitment Materials	07/Feb/2022	clean
2022_01_14_Johnsrude_119968	Recruitment Materials	14/Jan/2022	1
MTurk_Recruitment_C	Recruitment Materials	07/Feb/2022	clean
matrix_reasoning	Online Survey	10/Feb/2022	1
MentalRotationStimuli	Other Data Collection Instruments	10/Feb/2022	2
Qualtrics_debrief	Debriefing document	08/Mar/2022	1
11-backward_digit_span	Other Data Collection Instruments	08/Mar/2022	1
Additional sentence stimuli	Other Data Collection Instruments	08/Mar/2022	1
Headphone_check	Other Data Collection Instruments	08/Mar/2022	1
demographics_questionnaire	Online Survey	11/Mar/2022	4
LOI_Online_nonSONA_3	Implied Consent/Assent	16/Mar/2022	3
LOI_SONA_Online_4	Implied Consent/Assent	16/Mar/2022	4

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

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

Sincerely,

[redacted]

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix D: Materials for the secondary task

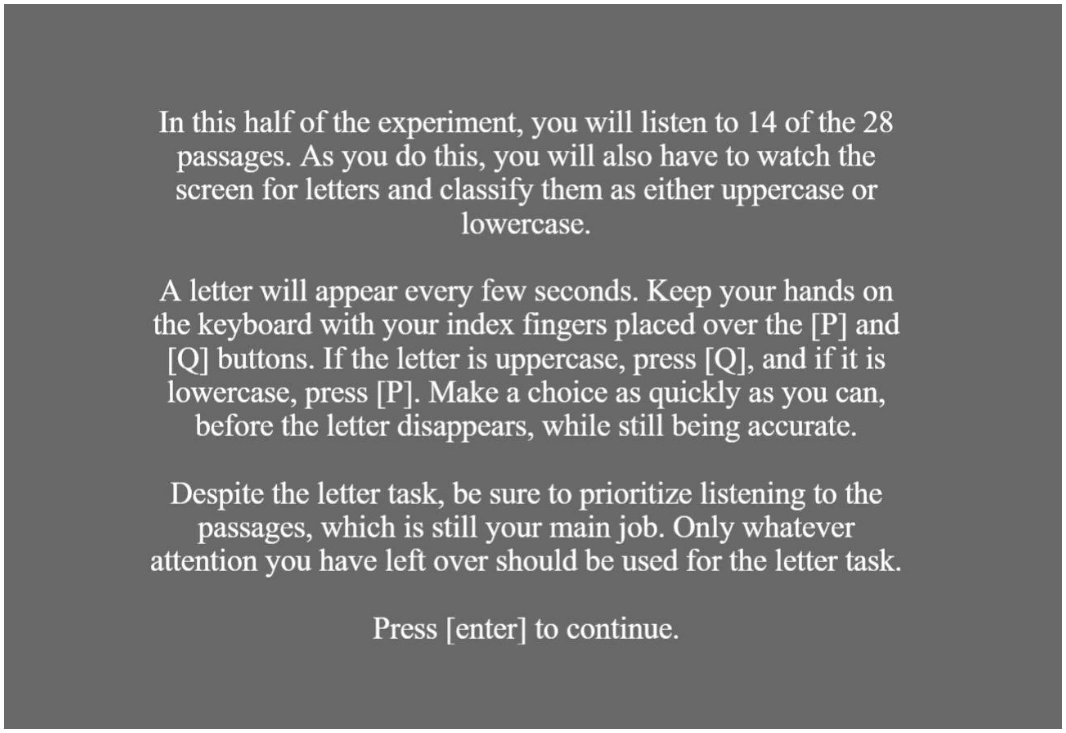

Figure D1: Instructions for the secondary task.

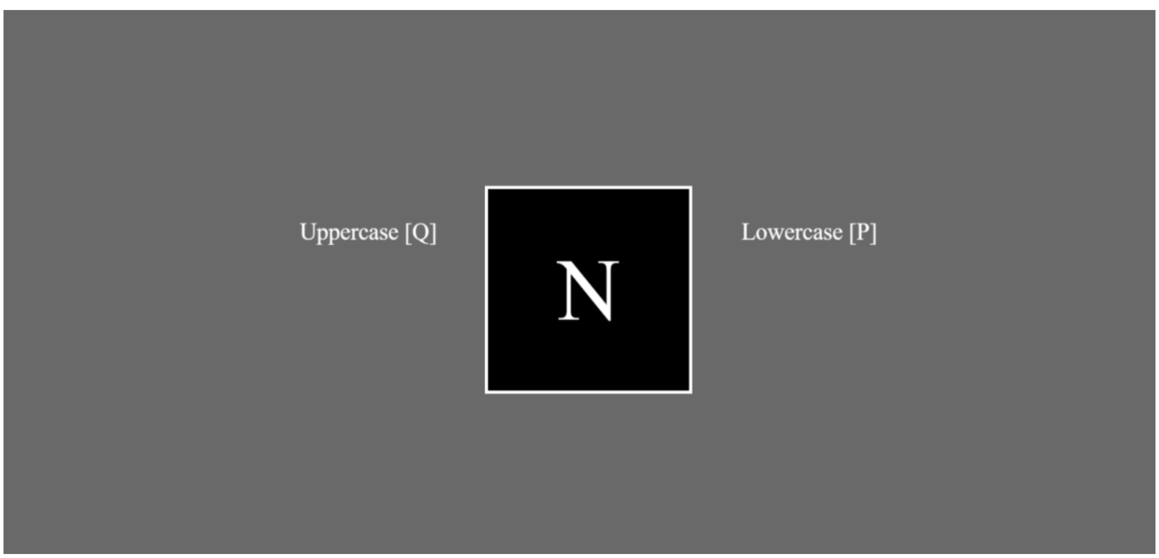

Figure D2: Prompt for the secondary task.

Appendix E: Materials for the effort rating scale

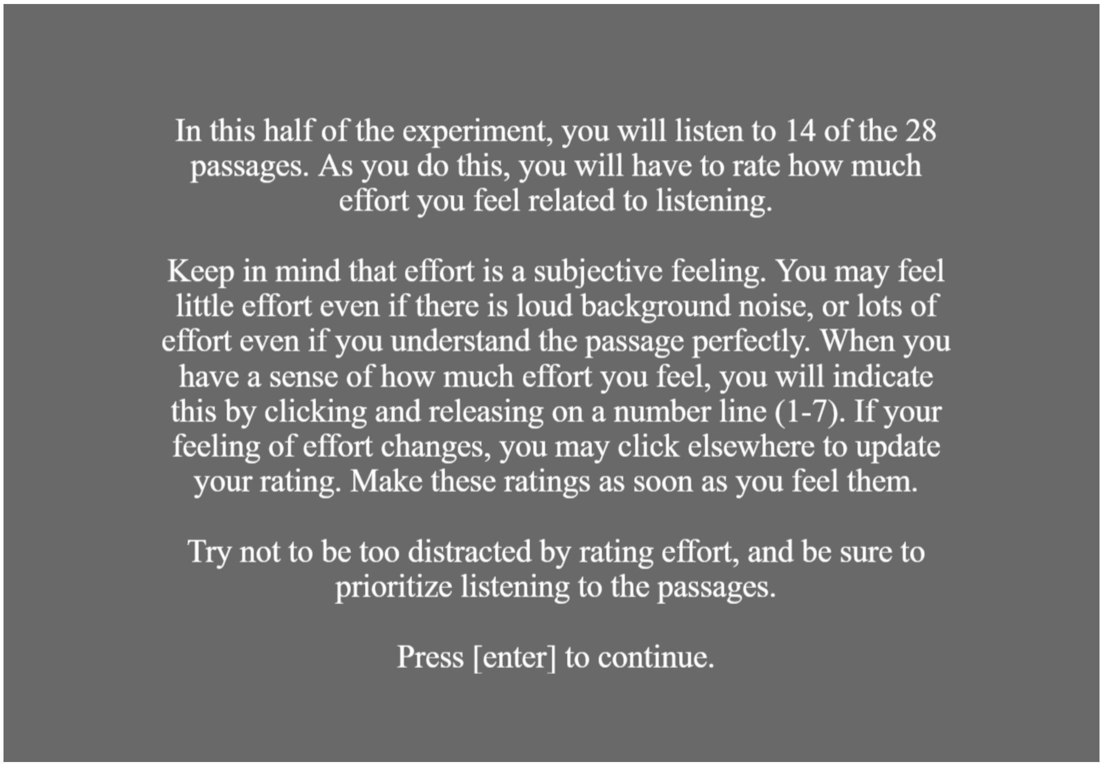

Figure E1: Instructions for the effort rating scale.

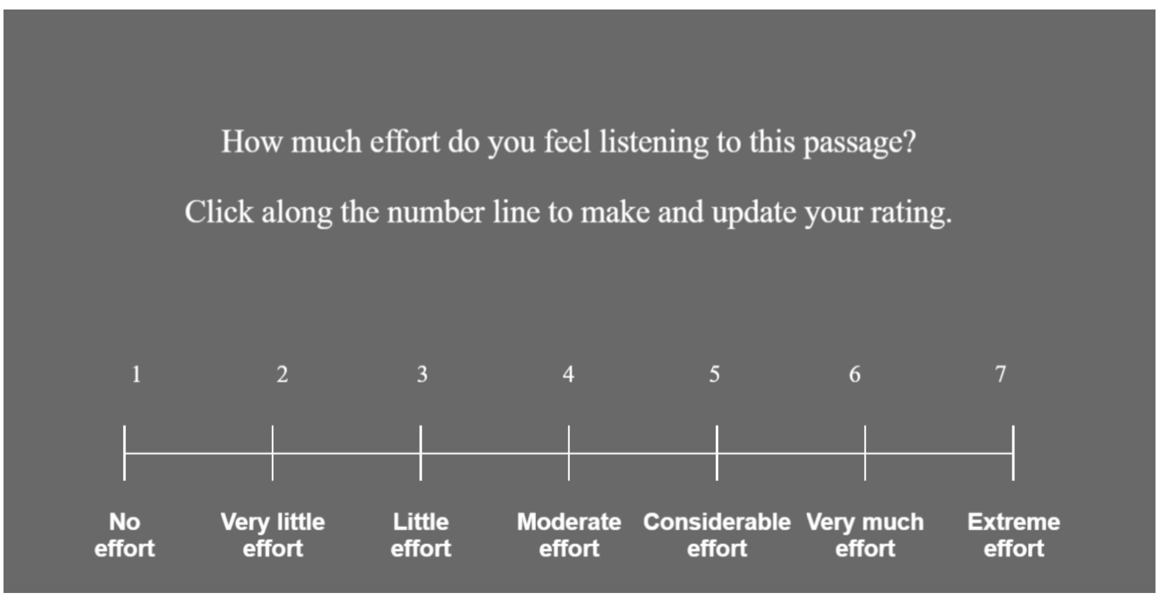

Figure E2: Prompt for the effort rating scale.

Appendix F: Secondary-task reaction time by letter position

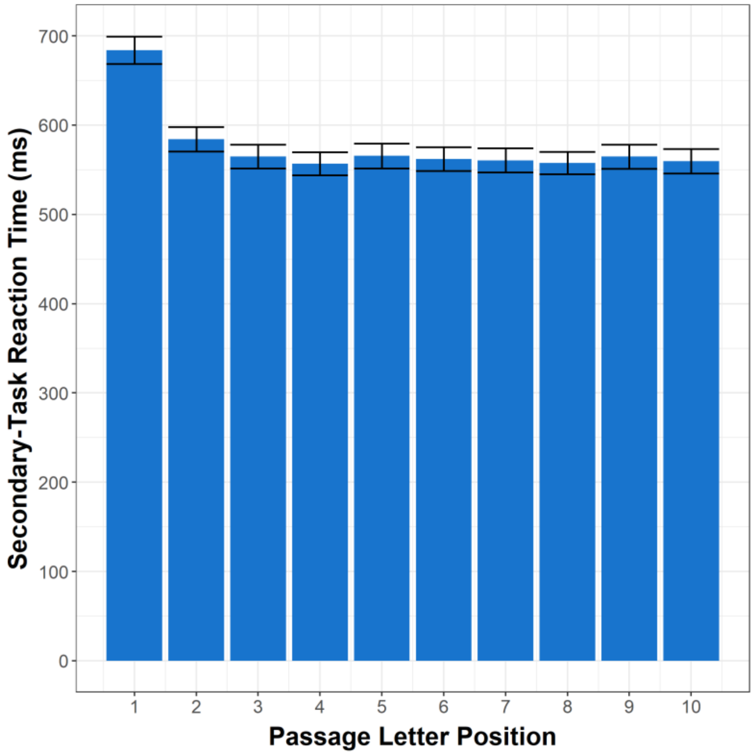

Figure F1: Bar heights represent the condition means and error bars represent the 95% confidence interval around the means. The first letter position had reaction times significantly slower than the other letter positions.

Appendix G: Analysis of comprehension

Figure G1 shows the effect of SNR on comprehension. Average comprehension was lower and varied less than gist understanding, ranging from 70% to 80% across SNRs. The final model of comprehension included random intercepts for participants but no random slopes. The effect of linear SNR on comprehension was not significant (*b* = 0.29, 95% CI = [-0.10, 0.67], β = 0.03, *t*[2628.14] = 1.45, *p* = .146), nor was the effect of negative exponential SNR (*b* = -1.31, 95% CI = [-3.26, 0.64], β = -0.03, *t*[2625.63] = -1.32, *p* = .187). The effect of RAPM score was significant (*b* = 0.79, 95% CI = [0.15, 1.44], β = 0.07, *t*[223.91] = 2.43, *p* = .016), with each one-point increase in RAPM score associated with an increase in gist understanding of just over 0.75 percentage-points on average (see Appendix H). The conditional R^2 of the final model was .11, meaning that the fixed and random effects accounted for 11% of the variance in comprehension.

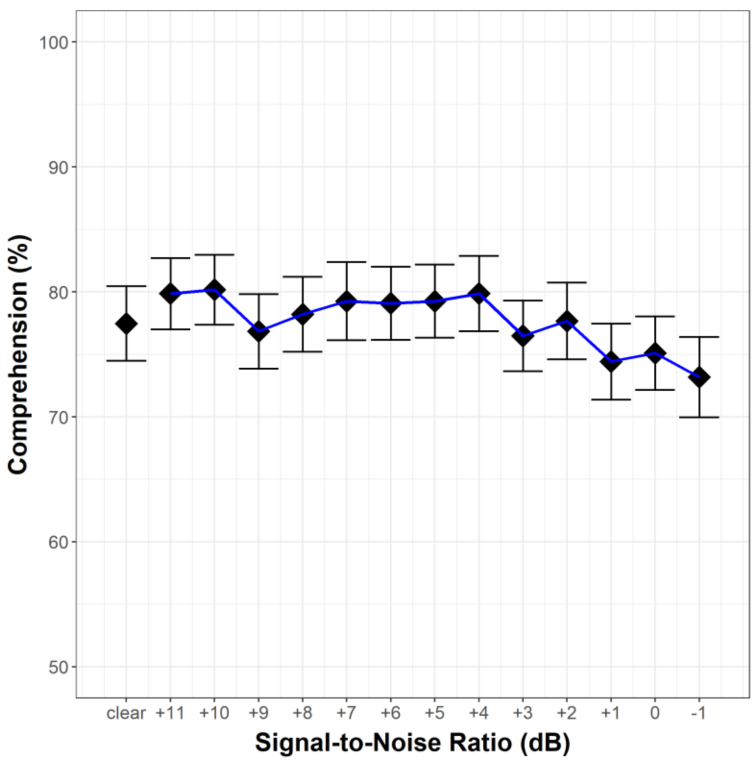

Figure G1: Comprehension questions were used to ensure that participants were attending to the passages and prioritized listening to them. Data points represent the condition means, error bars represent the 95% confidence intervals, and the blue lines connecting the means shows the trend.

Appendix H: Histogram of RAPM scores

Figure H1 shows the distribution of RAPM scores. RAPM scores, which could have a maximum value of 12, ranged from 0 to 11 (*M* = 6.12, *SD* = 2.69).

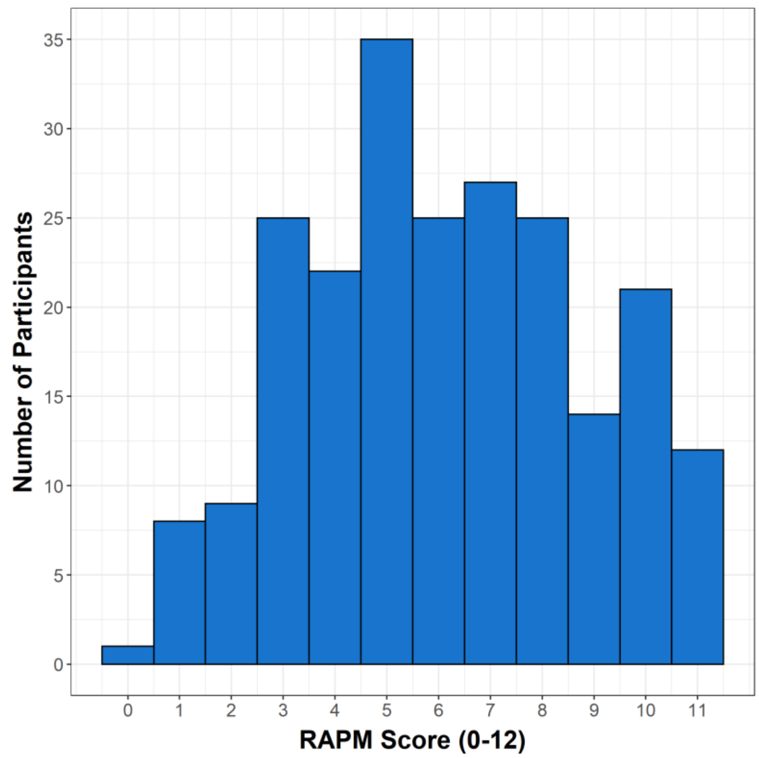

Figure H1: Bar heights represent the number of participants who achieved each RAPM score.

Curriculum Vitae

Joseph Rovetti

EDUCATION

M.Sc., Psychology – Western University Sep 2020 – Expected Aug 2022

- Advisor: Dr. Ingrid S. Johnsrude
- Thesis: "Listening effort: Separating the subjective from the objective"
- Academic average: 95.2%

Honours B.A., Psychology – Ryerson University Sep 2015 – Jun 2019

- Advisor: Dr. Frank A. Russo
- Thesis: "Functional near-infrared spectroscopy as a measure of listening effort in hearing-aided older adults"
- Cumulative GPA: 4.28/4.33

AWARDS AND SCHOLARSHIPS

- CIHR Vanier Canada Graduate Scholarship; ranked 4/178 ($150,000) 2022
  - 2022: "Understanding listening effort to improve the assessment and treatment of hearing loss"
- NSERC Canada Graduate Scholarship, Doctoral; ranked 3/198 ($105,000) 2022
  - 2022: declined for Vanier CGS
- Great Ideas for Teaching Award, First Place ($100) 2021
- Western Research Forum Best Oral Presentation Award, First Place ($30) 2021
- NSERC Canada Graduate Scholarship, Master's ($17,500)
  - 2020: "The role of motivation in the recruitment of the left inferior frontal gyrus under difficult listening conditions: A functional near-infrared spectroscopy study"
- Ontario Graduate Scholarship ($15,000) (x3) 2020 – 2022
  - 2022: declined for Vanier CGS
  - 2021: "Listening effort: Separating the subjective from the objective"
  - 2020: declined for CGS M
- Western Graduate Research Scholarship ($25,440 with renewals) 2020 – 2022
- Ryerson PGSA Research Symposium Best Poster Award, First Place ($50) 2019
- Governor General's Academic Silver Medal 2019
- Fourth-Year Psychology Award, Thesis Stream ($100) 2019
- CPA Certificate of Academic Excellence, Honours Thesis 2019

- Life Institute Student Award ($1000) 2019
- Irene Gammel & Jean-Paul Boudreau URA ($1000) 2019
- NSERC Undergraduate Student Research Award ($4500) (x3) 2017 – 2019
  - 2019: "Arousal effects on decision making in younger and older adults"
  - 2018: "Use of near infrared spectroscopy as an objective assessment of listening effort"
  - 2017: "Influence of short-term vocal training on auditory brainstem responses"
- Psychology Scholar's Circle Award (x4) 2016 – 2019
- Dean's List (x4) 2016 – 2019
- Ryerson Entrance Scholarship ($12,000 with renewals) 2015 – 2018

RESEARCH ITEMS

PEER-REVIEWED JOURNAL PAPERS

Counsell, A., **Rovetti, J.**, & Buchanan, E. (accepted). Psychometric evaluation of the *Students' Attitudes toward Statistics and Technology Scale (SASTSc). Statistics Education Research Journal*.

Rovetti, J., Copelli, F., & Russo, F.A. (2022). Audio and visual speech emotion activates the left pre-supplementary motor area. *Cognitive, Affective, & Behavioral Neuroscience, 22*, 291–303. doi:10.3758/s13415-021-00961-2

Copelli, F., **Rovetti, J.**, Ammirante, P., & Russo, F.A. (2022). Human mirror neuron responsivity to unimodal and multimodal presentations of action. *Experimental Brain Research, 240*, 537–548. doi:10.1007/s00221-021-06266-7

Rovetti, J., Goy, H., & Russo, F.A. (2021). Reduced semantic context and signal-to-noise ratio increase listening effort as measured using functional near-infrared spectroscopy. *Ear and Hearing, 43*, 836–848. doi:10.1097/AUD.0000000000001137

Ammirante, P., & **Rovetti, J.**, (2021). Bright vowels are favoured on weak beats in popular music lyrics. *Journal of New Music Research, 50*(5), 1–7. doi:10.1080/09298215.2021.1936076

Sullivan, M.D., Huang, R., **Rovetti, J.**, Sparrow, E., & Spaniol, J. (2021). Associations between phasic arousal and decisions under risk in younger and older adults. *Neurobiology of Aging*, 105, 262–271. doi:10.1016/j.neurobiolaging.2021.05.001

Rovetti, J., Goy, H., Nurgitz, R., & Russo, F.A. (2021). Comparing verbal working memory load in auditory and visual modalities using functional near-infrared spectroscopy. *Behavioural Brain Research, 402*, 1–10. doi:10.1016/j.bbr.2020.113102

Wood, E., **Rovetti, J.**, & Russo, F.A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition, 145*, 1–9. doi:10.1016/j.bandc.2020.105622

Rovetti, J., Goy, H., Pichora-Fuller, M.K., & Russo, F.A. (2019). Functional near-infrared spectroscopy as a measure of listening effort in older adults who use hearing aids. *Trends in Hearing, 23*, 1–22. doi:10.1177/2331216519886722

Rovetti, J., Behar, A., Abdoli, M., Copelli, F., & Russo, F.A. (2019). Listening effort in eateries. *Canadian Acoustics, 47*, 28. Retrieved from https://jcaa.caa-aca.ca/index.php/jcaa/article/view/3324/3209

OTHER PEER-REVIEWED PAPERS

Rovetti, J. (2020). Functional near-infrared spectroscopy: A clinical measure of listening effort? *Canadian Audiologist, 7*(6), 1–6. https://canadianaudiologist.ca/issue/volume-7-issue-6-2020/feature-6/

JOURNAL PAPERS IN PREPARATION

Hannah, J. A., **Rovetti, J.**, Van Hedger, S., & Johnsrude, I. S. (in preparation). Task effects on self-reported listening effort.

Hannah, J. A., Irsik, V. C., Herrmann, B., **Rovetti, J.**, & Johnsrude, I. S. (in preparation). A database of engaging stories for auditory research.

Rovetti, J., Popa, A., Ladowski, D., MacKinley, M., Dempster, K., Palaniyappan, L., & Johnsrude, I.S. (in preparation). Inviting hallucinatory percepts during speech-listening to detect cognitive changes in early psychosis.

Raynor, G.K., Johnsrude, I.S., Carlyon, R.P., **Rovetti J.**, & Deeks, J.M. (in preparation). Effects of aging on the time course of auditory perceptual organization of unresolved tone complexes.

Rovetti, J., Davis, M.H., Rodd, J., Hakyemez, H., Lee, S., Chau, A., & Johnsrude, I.S. (in preparation). The brain responds differently to acoustic and linguistic challenges to speech comprehension.

Russo, F.A., **Rovetti, J.**, & Wood, E.A. (in preparation). Reviewing the role of the motor system in the memory advantage for vocal melodies.

Rovetti, J., Sumantry, D., & Russo, F.A. (in preparation). Listening effort declines with exposure to nonnative-accented speech and predicts attitudes toward speakers.

CONFERENCE POSTER PRESENTATIONS

Rovetti, J., & Counsell, A. (June, 2021). *Measuring students' attitudes toward statistics and statistical software*. Poster presentation at the 82nd Annual Canadian Psychological Association Convention, online.

Rovetti, J., Copelli, F., Wood, E.A., Gilmore, S., & Russo, F.A. (November, 2020). *Speech emotion activates the pre-supplementary motor area*. Poster presentation at the 19th Annual Auditory Perception, Cognition, and Action Meeting, online.

Sullivan, M.D., Huang, R., **Rovetti, J.**, Sparrow, E., & Spaniol, J. (May, 2020). *The effect of phasic arousal on risky choice in younger and older adults*. Poster presentation at the 27th Annual Meeting of the Cognitive Neuroscience Society, online.

Rovetti, J., Sullivan, M.D., Huang, R., Sparrow, E., & Spaniol, J. (May, 2020). *Arousal effects on decision making in younger adults*. Poster presentation at the 1st Canadian Stress Research Summit, Toronto, ON [Conference cancelled].

Rovetti, J., Goy, H., & Russo, F.A. (November, 2019). *Is auditory working memory more demanding than visual?* Poster presentation at the 18th Annual Auditory Perception, Cognition, and Action Meeting, Montréal, QC.

Rovetti, J., Wood, E., Earle, E., & Russo, F.A. (November, 2019). *Vocal-motor interference disrupts the memory advantage for vocal melodies.* Poster presentation at the 11th Annual Psychology Graduate Students' Association Research Symposium, Toronto, ON.

Rovetti, J., Wood, E., Earle, E., & Russo, F.A. (November, 2019). *Vocal-motor interference disrupts the memory advantage for vocal melodies.* Poster presentation at the 15th Annual NeuroMusic Conference, Hamilton, ON.

Wood, E., **Rovetti, J.**, & Russo, F.A. (August, 2019). *Examining the role of the motor system in the vocal melody memory advantage.* Poster presentation at the Biennial Meeting of the Society for Music Perception and Cognition, New York City, NY.

Copelli, F., **Rovetti, J.**, Wood, E., Gilmore, S., & Russo, F.A. (June, 2019). *Internal simulation of speech emotion*. Poster presentation at the 29th Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Waterloo, ON.

Rovetti, J., Goy, H., Nurgitz, R., & Russo, F.A. (June, 2019). *The application of functional near-infrared spectroscopy to auditory research*. Poster presentation at the 29th Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Waterloo, ON.

Wood, E., **Rovetti, J.**, & Russo, F.A. (June, 2019). *Does motor system engagement contribute to the memory advantage for vocal melodies*? Poster presentation at the 29th Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Waterloo, ON.

Russo, F.A., **Rovetti, J.**, & Goy, H. (May, 2019). *Functional near-infrared spectroscopy as a measure of listening effort in hearing-aided older adults.* Poster presentation at the 4th Annual International Hearing Loss Conference, Niagara-on-the-Lake, ON.

Rovetti, J. (April, 2019). *Functional near-infrared spectroscopy as a measure of listening effort in hearing-aided older adults.* Poster presentation at the 11th Annual Psychology Undergraduate Thesis Poster Session at Ryerson University, Toronto, ON.

Rovetti, J., Goy, H., & Russo, F.A. (January, 2019). *Functional near-infrared spectroscopy as a measure of listening effort*. Poster presentation at the 14th Annual Toronto Rehabilitation Institute Research Day, Toronto, ON.

Nurgitz, R., **Rovetti, J.**, Goy, H., & Russo, F.A. (June, 2018). *Hearing difficulty and cognitive load: Using optical imaging to measure listening effort*. Poster presentation at the 29th Annual International Congress of Applied Psychology, Montréal, QC.

Rovetti, J., Nurgitz, R., Copelli, F., Goy, H., & Russo, F.A. (February, 2018). *Using functional near-infrared spectroscopy to measure cognitive load in the auditory domain*. Poster presentation at the 47th Annual Lake Ontario Visionary Establishment Conference, Niagara Falls, ON.

Rovetti, J., Nurgitz, R., Copelli, F., Goy, H., & Russo, F.A. (November, 2017). *Using functional optical neuroimaging to measure auditory working memory load*. Poster presentation at the 13th Annual NeuroMusic Conference, Hamilton, ON.

CONFERENCE ORAL PRESENTATIONS

Rovetti, J. (May, 2021). *Teaching registered reports for psychological science*. Oral presentation at the Annual Western Own Your Future May Conference on Teaching, online.

Russo, F. A., Wood, E. A., **Rovetti, J.**, & Kovacek, K. (June, 2021). *The role of sensorimotor simulation in the memory advantage for vocal melodies.* Oral presentation at the 30th Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, online.

Rovetti, J., Sumantry, D., & Russo, F.A. (March, 2021). *Listening effort declines with exposure to nonnative-accented speech and predicts attitude toward the speaker*. Oral presentation at the 35th Annual Western Research Forum, online.

Rovetti, J., Goy, H., & Russo, F.A. (May, 2020). *How does semantic context affect listening effort?* Oral presentation at Applied Cognitive Neuroscience at Ryerson University, online.

Copelli, F., **Rovetti, J.**, Ammirante, P., & Russo, F.A. (May, 2020). *EEG analysis of the mirror neuron system in humans*. Oral presentation at Applied Cognitive Neuroscience at Ryerson University, online.

Rovetti, J., Goy, H., & Russo, F.A. (May, 2020). *What is listening effort and what factors influence it?* Oral presentation at the 4th Meeting of the Toronto Auditory Research Group, Toronto, ON [Conference cancelled].

Copelli, F., **Rovetti, J.**, Ammirante, P., & Russo, F.A. (May, 2020). *Investigating involvement of the action-observation network in audition using EEG*. Oral presentation at the 4th Meeting of the Toronto Auditory Research Group, Toronto, ON [Conference cancelled].

Counsell, A., & **Rovetti, J.** (May, 2020). *Measuring students' attitudes toward using statistical software*. Oral presentation at the 81st Annual Canadian Psychological Association Convention, Montréal, QC [Conference cancelled].

Russo, F.A., **Rovetti, J.**, & Goy, H. (March, 2020). *Using fNIRS to measure listening effort* [Presentation cancelled]. Oral presentation at the 47th Annual Scientific and Technology Conference of the American Auditory Society, Scottsdale, AZ.

Behar, A., **Rovetti, J.**, Abdoli, M., Copelli, F., & Russo, F.A. (October, 2019). *Listening effort in eateries.* Oral presentation at the annual Acoustics Week in Canada, Edmonton, AB.

Rovetti, J. (May, 2019). *Functional near-infrared spectroscopy as a measure of listening effort in hearing-aided older adults.* Oral presentation at the 49th Annual Ontario Psychology Undergraduate Thesis Conference, Waterloo, ON.

Rovetti, J., Goy, H., & Russo, F.A. (January, 2019). *Functional near-infrared spectroscopy as a measure of listening effort*. Oral presentation at the 14th Annual Toronto Rehabilitation Institute Research Day, Toronto, ON.

OTHER PRESENTATIONS

Rovetti, J. (November, 2021). *Equivalence testing and implementation in R*. Guest lecture for PSYCHOL 9041 at Western University.

Rovetti, J. (October, 2021). *Power and power analyses in R: A primer*. Workshop for the Department of Psychology at Western University.

Rovetti, J. (November, 2020). *Equivalence testing in R: A primer*. Workshop for the Department of Psychology at Western University.

Rovetti, J. (March, 2019). *Listening effort, and how it affects older adults with hearing loss*. Oral presentation at the LIFE Institute Partners in Learning Event at Ryerson University.

POSITIONS HELD

ACADEMIC POSITIONS

School of Graduate and Postdoctoral Studies – Western University

Graduate Student/Teaching Assistant Sep 2020 – Present

Courses TA'd:

- Introduction to Statistics Using R (PSYCHOL 9041; graduate) Sep 2021 – Dec 2021
- Introduction to Psychology (PSYCHOL 1000; undergraduate) Sep 2020 – Dec 2020

RESEARCH POSITIONS

Cognitive Neuroscience of Communication and Hearing Lab – Western University
Supervisor: Dr. Ingrid Johnsrude
Graduate Research Assistant — Aug 2020 – Present

Multisensory Integration in Virtual Environments Lab – Toronto Rehab
Supervisor: Dr. Jennifer Campos
Research Associate — Sep 2019 – Aug 2020

Research Methods and Statistics Lab – Ryerson University
Supervisor: Dr. Alyssa Counsell
Assistant Data Scientist — Sep 2019 – Aug 2020

Memory and Decision Processes Lab – Ryerson University
Supervisor: Dr. Julia Spaniol
NSERC USRA Research Assistant — Apr 2019 – Sep 2019

Sonova Holding AG (Swiss and Canadian locations)
Research Consultant — Jun 2018 – Mar 2020

Science of Music, Auditory Research, and Technology Lab – Ryerson University
Supervisor: Dr. Frank A. Russo
Research Associate — Sep 2019 – Aug 2020
NSERC USRA Research Assistant/Lab Manager — May 2018 – Aug 2018
NSERC USRA Research Assistant — May 2017 – Aug 2017
Research Assistant — Jan 2017 – Aug 2019

Cognitive Aging Lab – Ryerson University
Supervisor: Dr. Lixia Yang
Research Intern — Oct 2016 – Mar 2017

Cognitive Neuroscience Lab – Ryerson University
Supervisor: Dr. Tisha Ornstein
Research Assistant — Aug 2017