

# Measuring the Dual-Task Costs of Audiovisual Speech Processing Across Levels of Background Noise

Violet A. Brown

Department of Psychological & Brain Sciences, Washington University in St. Louis

Successful communication requires that listeners not only identify speech, but do so while maintaining performance on other tasks, like remembering what a conversational partner said or paying attention while driving. This set of four experiments systematically evaluated how audiovisual speech—which reliably improves speech intelligibility—affects dual-task costs during speech perception (i.e., one facet of *listening effort*). Results indicated that audiovisual speech *reduces* dual-task costs in difficult listening conditions (those in which visual cues substantially benefit intelligibility) but may actually *increase* costs in easy conditions—a pattern of results that was internally replicated multiple times. This study also presents a novel dual-task paradigm specifically designed to facilitate conducting dual-task research remotely. Given the novelty of the task, this study includes psychometric experiments that establish positive and negative control, assess convergent validity, measure task sensitivity relative to a commonly used dual-task paradigm, and generate performance curves across a range of listening conditions. Thus, in addition to evaluating the effects of audiovisual speech across a wide range of background noise levels, this study enables other researchers to address theoretical questions related to the cognitive mechanisms supporting speech processing beyond the specific issues addressed here and without being limited to in-person research.

## Public Significance Statement

Previous work has shown that seeing the talker in addition to hearing their voice improves the listener's ability to understand spoken language. This work shows that despite this benefit, seeing the talker may make it more difficult to multitask, but only in easy listening conditions in which seeing the talker is not necessary to understand speech. When the background noise is loud, however, seeing the talker both improves speech identification and helps the listener perform simultaneous tasks. Finally, this project includes important measurement work and large samples of online participants to ensure that the findings are robust and the task is maximally useful to other researchers studying related phenomena.


**Keywords:** audiovisual speech, listening effort, dual-task costs, measurement, validation

**Supplemental materials:** <https://doi.org/10.1037/xge0001826.sup>

Understanding spoken language requires that listeners extract meaning from a complex and rapidly changing acoustic signal. Despite the remarkable ease and efficiency with which many listeners comprehend continuous speech, parsing this bottom-up input is not a trivial task. For example, speech often occurs in the presence of background noise, such as the hum of an air vent, the babble of cafeteria noise, or a simultaneous conversation. Regardless of the source or type of noise, the target speech must be segregated from simultaneous acoustic inputs. This poses a challenge not only on a

sensory level, whereby extraneous noise renders the target signal inaudible as a result of spectral overlap, but also on a cognitive level, as the listener may have difficulty maintaining focus on the target speech in the presence of simultaneous conversations or other forms of background noise (e.g., Freyman et al., 1999). All of this must be accomplished in the face of substantial intraperson (i.e., coarticulation) as well as interperson variability in speech production. The signal then disappears, leaving only a brief trace in echoic memory (see, e.g., Buchsbaum et al., 2005). Thus, although comprehending

Joseph Toscano served as action editor.

Violet A. Brown  <https://orcid.org/0000-0001-5310-6499>

All data, materials, code, and preregistrations can be accessed on the Open Science Framework at <https://osf.io/pqj8h/>. Some of the data reported in this article were presented at the 64th Annual Meeting of the Psychonomic Society. This work was supported by Washington University in St. Louis. The author is grateful to Kristin Van Engen, Julia Strand, and Mitchell Sommers for helpful

discussions and feedback on an earlier draft of the article and to Jess Sims and Leanne Hainsby for their support during the writing of this article.

Violet A. Brown played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, and writing—review and editing.

Correspondence concerning this article should be addressed to Violet A. Brown. Email: [violetbrown@carleton.edu](mailto:violetbrown@carleton.edu)

continuous speech appears to proceed effortlessly for most people (Rönnerberg et al., 2008), understanding spoken language is an extraordinarily complex task.

The cognitive cost associated with processing speech is often referred to as *listening effort* and manifests as poorer performance on simultaneous tasks (i.e., impaired ability to multitask; Desjardins & Doherty, 2014), poorer recall of what was heard (Rabbitt, 1968), and increased physiological arousal (e.g., greater pupil dilation; McGarrigle et al., 2017). In listening conditions that require high levels of listening effort, listeners rely on several cues to facilitate speech understanding. For example, when speech is presented in background noise and intelligibility falters, listeners make use of visual cues provided by the talking face to understand the spoken message (Erber, 1972; Sumby & Pollack, 1954). However, despite the clear intelligibility benefits of audiovisual speech, the literature is mixed as to whether visual speech cues incur additional processing costs beyond audio-only speech. Indeed, there is evidence that seeing the talker may increase (Brown & Strand, 2019a), decrease (Sommers & Phelps, 2016), or not affect listening effort (Keidser et al., 2015) relative to hearing alone.

The first objective of this study was to assess whether and how seeing the talker affects listening effort. Although there are many methods for measuring listening effort (see, e.g., Strand et al., 2018; Strand, Ray, et al., 2020), this study specifically focused on the dual-task costs associated with listening; these costs were therefore used not only as a measure of multitasking ability but also as a measure of this particular feature of the multidimensional construct of listening effort. To enable more direct comparison with prior work (e.g., Brown & Strand, 2019a), the present study used isolated spoken words as stimuli, but the novel paradigm reported here was designed with the intention that it could be extended to sentence-length materials as well. The remaining goals of this study were methodological rather than theoretical in nature. The listening effort literature suffers from insufficient attention to task validation, and no listening effort measure has been validated for research conducted remotely. The second objective of this study was therefore to validate a novel dual-task paradigm that was developed to measure the costs associated with processing audio-only and audiovisual speech in experiments conducted remotely or in-lab.

## Background

### Listening Effort

Despite the subjective ease with which most individuals process speech, converting a noisy acoustic input into meaningful mental representations does not occur automatically; that is, identifying speech requires cognitive resources that exist in finite amounts (Kahneman, 1973). More challenging listening environments necessitate the recruitment of additional cognitive resources, which diverts resources away from simultaneous tasks and therefore impairs the listener's ability to quickly and accurately complete additional tasks. Indeed, a substantial body of evidence suggests that when a secondary cognitive task is performed simultaneously with a speech identification task, performance on the secondary task is impaired as the difficulty of the speech task increases (e.g., Brown & Strand, 2019b; Rabbitt, 1968; Zekveld & Kramer, 2014). Taken together, these results suggest that despite the perceived automaticity of speech comprehension, this process does not occur resource-free.

The Ease of Language Understanding model (Rönnerberg et al., 2008, 2010)—a framework for considering the perceptual and cognitive interactions underlying speech identification in noise—is particularly relevant to the concept of listening effort. According to this model, speech processing in ideal listening conditions proceeds automatically. However, in suboptimal conditions that produce mismatches between the incoming acoustic signal and phonological and lexical representations stored in memory (whether introduced by background noise, hearing loss, or some other factor), cognitive resources must be recruited to reconstruct the impoverished sensory input and infer meaning. The Ease of Language Understanding model can therefore explain why background noise (as well as hearing loss, reverberation, and so on) increases listening effort as measured by dual-task costs, pupil dilation, and many other measures: Simultaneous overlapping frequency bands in the background noise render portions of the target speech inaudible, so listeners must recruit additional cognitive resources to reconstruct the impoverished input (i.e., they must exert listening effort).

### Audiovisual Speech

To overcome the challenges of a noisy acoustic input, listeners make use of several cues both internal and external to the spoken message. Of particular benefit for successful communication in face-to-face settings are the phonetic and temporal cues provided by the talking face. Indeed, one of the most robust findings in the speech perception literature is that seeing as well as hearing the talker improves speech identification in noise relative to hearing alone (Erber, 1972; Sumby & Pollack, 1954). This “audiovisual benefit” has been observed for syllables, words, and sentences (Sommers et al., 2005); in individuals with normal hearing and those with hearing loss (Tye-Murray et al., 2007); in cochlear implant users (Kaiser et al., 2003); and across the adult lifespan (Tye-Murray et al., 2016).

The benefit of audiovisual speech derives from the fact that the visual signal provides both complementary and redundant phonetic and temporal cues to the auditory signal (Campbell, 2008; Grant & Walden, 1996). The two modalities are complementary in that they differ in the phonetic cues they most readily provide: Voicing and manner of articulation are easily obtained from the auditory signal but difficult to extract from the visual signal, whereas place of articulation is visually apparent for some phonemes (e.g., distinguishing between bilabials and alveolar stops) but may be lost in noise (Walden et al., 1974). For example, the word “bag” may be mistaken for “tag” in noisy audio-only conditions, but seeing the talker's face easily disambiguates the two words because the onsets have different places of articulation. The two modalities are redundant in that some of the same phonetic and temporal cues are readily provided by both modalities. For example, the area of the opening between the lips is correlated with the acoustic amplitude envelope of the speech (Grant & Seitz, 2000), suggesting that the visual signal (in addition to the auditory signal) can provide information about the onset of the speech as well as other salient moments in the speech stream (e.g., pauses, moments of peak amplitude; Bernstein et al., 2004; Tye-Murray et al., 2011).

### Audiovisual Speech and Listening Effort

In addition to the large body of work demonstrating that seeing the talker improves speech intelligibility in noise relative to hearing

alone (e.g., Erber, 1975; Sumbly & Pollack, 1954), some studies using recall paradigms—in which participants listen to speech in noise and are later asked to recall what they heard—have shown that audiovisual speech reduces listening effort (Mishra et al., 2013; Sommers & Phelps, 2016). This finding is likely driven by two factors: Not only does the phonetic information provided by the face reduce the number of mismatches between the input and lexical representations—that is, phonetic cues mitigate the resource-intensive process of *lexical competition* (Kuchinsky et al., 2013; Wagner et al., 2016)—but visual cues may also reduce the attentional demands placed on auditory speech processing by automatically directing attention to salient acoustic features (Tye-Murray et al., 2011), thereby freeing resources that would otherwise be devoted to the speech task.

Other studies, however, have shown that audiovisual speech incurs additional processing costs relative to audio-only speech (Brown & Strand, 2019a; Fraser et al., 2010), perhaps because integrating the two unimodal inputs into a unified percept is a resource-intensive process (Alsius et al., 2005) or because simultaneously monitoring two channels is cognitively demanding. Still other work has found that audio-only and audiovisual speech do not differ in their processing costs (Keidser et al., 2015). One explanation for these discrepant findings that has some support in the literature is that the visual signal differentially affects effort depending on the difficulty of the listening task (Brown & Strand, 2019a; Mishra et al., 2013). When the listening conditions are difficult and the visual signal substantially improves intelligibility relative to audio-only conditions, the benefit of reduced lexical competition outweighs any processing costs associated with integrating the inputs from the two modalities, leading to a *decrease* in effort. When the listening conditions are easy, however, the extraneous visual signal incurs a processing cost without the benefit of reduced lexical competition, leading to an *increase* in listening effort. Finally, when these costs and benefits are in approximate equilibrium, no difference in effort is observed.

## Measuring Listening Effort

In addition to the explanation that the various mechanisms underlying audiovisual speech processing may differ in their cognitive demands (i.e., integrating the unimodal inputs may be resource-intensive, whereas temporal correspondence and weakened lexical competition may reduce listening effort), the discrepant findings may also be attributable to inconsistencies in how listening effort is measured (see Alhanbali et al., 2019; McGarrigle et al., 2014; Rudner et al., 2012; Strand et al., 2018).

Listening effort tasks are typically divided into three broad classes: subjective self-report (Borg, 1990; Johnson et al., 2015; McGarrigle et al., 2021); physiological, including heart rate variability, skin conductance (Mackersie & Cones, 2011; Seeman & Sims, 2015), and pupillometry (Beatty, 1982; Koelewijn et al., 2012; McGarrigle et al., 2017; Ohlenforst et al., 2018; Zekveld et al., 2010); and behavioral measures of listening effort, including dual-task (Brown et al., 2020; Pals et al., 2013, 2015; Strand, Brown, & Barbour, 2020) and recall paradigms (McCoy et al., 2005; Ng et al., 2013; Pichora-Fuller et al., 1995; Sommers & Phelps, 2016). One study found that across these three broad classes, more than two dozen different tasks have been used to measure listening effort (Strand et al., 2018), a number that

has increased substantially since that research was conducted more than 7 years prior to the publishing of this article.

Although an exceedingly large number of tasks have been used to assess listening effort (Strand et al., 2018), a review of the literature suggests that variations of dual-task paradigms are used more frequently than any other category of listening effort measure (see Gagné et al., 2017, for a review). It is unclear why these tasks tend to be the default in the listening effort literature, but this tendency may be driven by the fact that dual-task paradigms are relatively easy to implement (compared to, e.g., physiological measures), and the primary and secondary tasks can be modified to accommodate a wide range of speech stimuli and listening environments (e.g., audio-only words, audiovisual passages). Another explanation for their widespread use is that dual-task paradigms have a high degree of ecological validity; indeed, multitasking situations are common in everyday life (e.g., talking on the phone while driving; Strayer & Johnston, 2001), and dual-task paradigms reflect these real-world circumstances. Finally, dual-task paradigms have theoretical roots outside of the listening effort literature (Kahneman, 1973; see below), making them applicable to other domains within cognitive psychology. Given their prevalence in the literature, their flexibility and ease of implementation, and their ecological validity, the present study focuses on dual-task costs, which are expected to increase as listeners exert more effort.

## Dual-Task Paradigms

### *Multiple Resource Theory*

Although the construct of listening effort relies on the assumption that humans have a limited pool of cognitive resources that can be allocated to various tasks, this need not imply that there exists a single undifferentiated pool of resources. Indeed, Kahneman himself acknowledged that changing the structure of the secondary task affects the degree to which it interferes with the primary task. This observation suggests the existence of independent pools of specific resources, whereby simultaneous tasks only interfere with one another to the extent that they compete for the same pools of resources.

This theory has been expanded into *multiple resource theory*, which postulates that in addition to competing for a general pool of resources, simultaneous tasks compete for “satellite” pools of resources that are dedicated to complete specific categories of tasks (Fleming et al., 2024; Isreal et al., 1980; Navon & Gopher, 1979; Wickens, 1981). There is evidence for at least four independent pools of resources representing dichotomies of information processing: stage of processing (e.g., perception/cognition vs. response), code of processing (e.g., spatial vs. verbal/linguistic), perceptual modality (which is assumed to be nested within perception and unrelated to cognition or response; e.g., auditory vs. visual), and visual channel (e.g., focal vs. ambient vision; Wickens, 2008). Tasks that rely on the same pool of resources (e.g., two spatial tasks) are expected to provide more interference than those that rely on separate resource pools (e.g., a spatial task and a verbal task) because the pools are independent, a claim that has been supported in the listening effort literature (Picou & Ricketts, 2014) as well as in the dual-task literature more generally (Isreal et al., 1980). Thus, according to multiple resource theory, provided that the secondary task has resource overlap with the primary speech task, almost any combination of speech and

secondary tasks may be used to assess the dual-task costs associated with speech processing.

### A Novel Dual-Task Paradigm

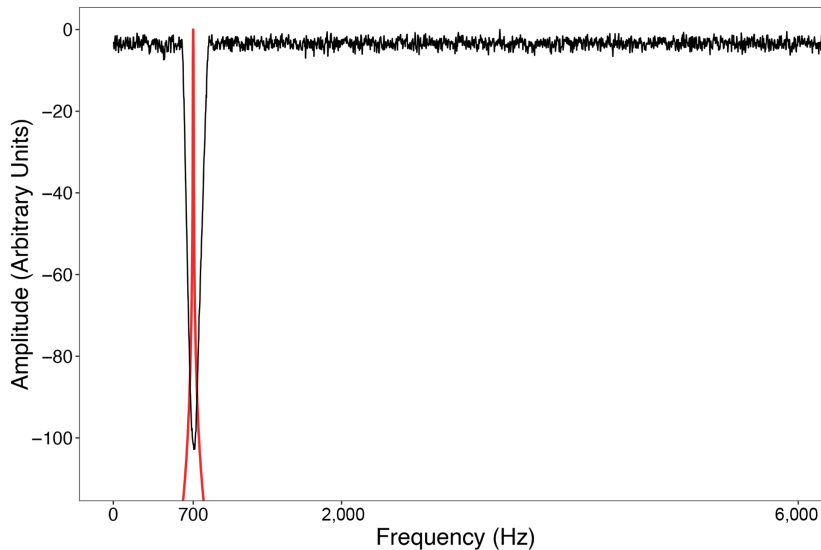
A clear benefit of dual-task paradigms is that they assess the difficulty of the listening situation in real time and reveal the tangible behavioral consequences of processing speech while simultaneously performing other tasks (e.g., driving, taking notes). However, a challenge of using dual-task paradigms to assess the cognitive costs associated with audiovisual speech processing is that if the secondary task occurs in the auditory or visual modality, any changes in performance on that task between audio-only and audiovisual speech conditions may be due to *sensory* (or *peripheral*) interference rather than *cognitive* interference. For example, a participant may respond more slowly to a flash of light in audiovisual conditions not because they expended greater effort to process audiovisual speech, but because the video of the talker reduced the visibility of the flash of light. In other words, physical constraints on the processing system rather than changes in resource allocation might account for differences between conditions (Wickens, 1981). To overcome this issue and ensure that changes in secondary task performance are attributable to cognitive effort rather than sensory interference, prior work has used secondary tasks in the tactile modality—in which the secondary task involves classifying pulses presented to a finger as short, medium, or long in duration—when assessing dual-task costs for audiovisual speech (e.g., Brown & Strand, 2019a; Fraser et al., 2010; Gosselin & Gagné, 2011a). Although this approach has revealed reliable differences in dual-task costs as a function of the modality in which the speech is presented, this method requires custom-built hardware, making it (a) difficult and time-consuming to implement, (b) nearly impossible for other

labs to directly replicate, and (c) incompatible with research conducted remotely in an online setting.

For the present study, I developed a novel dual-task paradigm designed for use in both audio-only and audiovisual speech perception experiments. In this task, participants identified speech in background noise while simultaneously classifying tones as short, medium, or long in duration (an auditory analogue to the vibrotactile task used in Brown & Strand, 2019a), and response times to the tone classification task were taken as a measure of the dual-task costs associated with identifying the speech. Crucially, the tone was always presented at an audible level, and the frequency of the tone was distinct from the frequencies present in the background noise to ensure that changes in background noise level do not affect audibility of the tone (Figure 1). Although the present study validated this approach using single-word stimuli, the paradigm was explicitly designed to be extensible to speech materials consisting of sentences or narrative passages. Thus, not only does this paradigm enable researchers to answer theoretical questions related to multimodal speech processing in the presence of background noise, but it also allows them to assess dual-task costs at multiple levels of linguistic analysis and various levels and types of contextual constraint.

Given increasing interest in online research, this task was designed to enable researchers to assess the dual-task costs associated with audiovisual speech processing remotely. In addition to being robust to pandemic-related campus closures, online research facilitates the implementation of highly powered studies in a relatively short time frame, increases opportunities for research participation, and increases the diversity of participant pools (see, e.g., Henrich et al., 2010). As such, results from online studies are likely to be more generalizable than those from studies in which data were collected in highly controlled lab settings. Additionally, several studies across domains of psychology have shown consistent patterns of results across online and in-lab studies (Basu Mallick et al., 2015; Sheehan, 2018; Slote &

**Figure 1**  
*Frequency-Domain Plot of the Notched Noise (Black, Thinner Line) and 700 Hz Tone (Red or Gray, Thicker Line)*



*Note.* The frequency of the tone has been removed from the background noise. See the online article for the color version of this figure.



Strand, 2016), and online platforms have been shown to collect reliable response time data across web browsers, operating systems, and hardware (Anwyl-Irvine et al., 2021).

An additional benefit of using this paradigm beyond its applicability to a wide range of speech stimuli and its ability to be implemented remotely is that the secondary task is in the same modality as the primary task, meaning that according to multiple resource theory, the two tasks compete for the general pool of resources as well as a satellite pool devoted to auditory processing. It was therefore expected that this tone classification task would be more sensitive to changes in the difficulty of the speech task relative to tasks in other modalities (e.g., a secondary task in the tactile modality; Brown & Strand, 2019a); indeed, previous work has demonstrated that even after controlling for peripheral interference, tasks in the same sensory modality compete for resources to a greater extent than those in different modalities (see Wickens, 1981). This potential increase in task sensitivity is beneficial in terms of both statistical power and applicability to research areas in which small effect sizes are common or expected.

### The Present Study

Seeing the talker in addition to hearing their voice improves speech intelligibility in noise, but it is unclear whether and how it affects dual-task costs. The present study fills these gaps in the literature using highly powered designs and robust statistical techniques. This study includes a conceptual replication of a previous dual-task experiment demonstrating that processing audiovisual relative to audio-only words incurs additional dual-task costs in easy but not hard levels of background noise (Brown & Strand, 2019a). By replicating and then building upon previous work by including a wide range of signal-to-noise ratios (SNRs), this project reveals the real-time cognitive costs of processing spoken language in a form in which it is frequently encountered: audiovisually and in the presence of background noise.

A widespread issue in the listening effort literature (and in many areas within and beyond psycholinguistics) is the lack of attention to the validity of the measurement tools used to evaluate psychological constructs (Flake & Fried, 2020; Strand, Ray, et al., 2020). Without sufficient attention to validity, researchers can only draw conclusions about specific tasks in specific experimental contexts. Thus, in addition to addressing the theoretical question of how audiovisual speech affects dual-task costs across levels of background noise, this study also assessed the validity and psychometric properties of the novel dual-task paradigm. Specifically, *Experiment 1* (audio-only and audiovisual) established positive control by ensuring that the measure was sensitive to changes in SNR when speech was present. *Experiment 2* (audio-only) established negative control by demonstrating that the measure was *not* sensitive to changes in the level of noise when speech was *not* present. Together, Experiments 1 and 2 ensured that the measure was truly sensitive to the challenges of listening to speech in noise rather than either noise-induced cognitive interference (see Brown & Strand, 2019b; Colle & Welsh, 1976; Weisz & Schlittmeier, 2006) or inaudibility of the tone as the level of the background noise increased. *Experiment 3* (audio-only) correlated performance on the novel tone task with performance on another widely used dual-task paradigm (see Brown et al., 2020; Brown & Strand, 2019b; Picou & Ricketts, 2014; Sarampalis et al., 2009) to assess convergent validity. Experiment 3 also compared the

sensitivities of the two tasks to changes in SNR. *Experiment 4* (audio-only and audiovisual) obtained psychometric information about the task in the form of performance curves across a wide range of SNRs. These validation experiments were conducted and administered online to provide evidence supporting the validity of this dual-task measure specifically for online research.

### General Method and Analysis Plan

Table 1 outlines the four experiments in this study and provides a brief description of each.

All procedures were approved by the Washington University in St. Louis Institutional Review Board, experiments were conducted online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), and the method and analysis plan were preregistered via the Open Science Framework (see below for details). Participants were recruited through the Washington University in St. Louis Department of Psychological & Brain Sciences subject pool and through Prolific (<https://www.prolific.com>; Palan & Schitter, 2018). All participants had self-reported normal hearing and normal or corrected-to-normal vision. Participants were required to complete the task using Google Chrome due to constraints on the presentation of certain file types online. Stimuli were unique across experiments, so participants were eligible to participate in multiple experiments. All data, materials, and code are available at <https://osf.io/pqj8h/>.

### Stimuli

The tone stimuli consisted of a 700-Hz sine wave tone created using Audacity's "Filter EQ Curve" algorithm (Audacity Version 2.3.3). On any given trial, the tone was either 90 ms (short), 150 ms (medium), or 250 ms (long) in length. When speech was present (Experiments 1, 3, and 4), the tones occurred at four points relative to the onset of the speech in 190 ms increments (see below). Unless otherwise noted, the level of the tone was set to  $-28$  dB root mean square (RMS), and speech stimuli were set to  $-24$  dB RMS.<sup>1</sup>

The background noise consisted of white noise<sup>2</sup> with a narrow frequency band surrounding 700 Hz removed (ranging from approximately 620 to 800 Hz), resulting in limited spectral overlap between the tone and the background noise (see Figure 1). Although this setup means that there was spectral overlap between the speech and tone (i.e., a narrow band surrounding 700 Hz was *not* removed from the speech), given that all words appeared in all conditions, this spectral overlap cannot systematically affect results. However, to further address this concern, the tone always occurred at the same point in a given word to ensure that the relationship between the speech and tone was constant across conditions. The same type of background noise (referred to as "notched") and tones were used in all experiments, but the SNRs varied by experiment. Tone, noise, and speech files were mixed using ffmpeg (Version 5.0; Tomar, 2006). A pilot study revealed that speech identification accuracy was slightly better in the notched noise than in the standard noise.

<sup>1</sup> Note that it is not possible to determine the exact sound pressure level at which stimuli were delivered to participants because they were instructed to adjust their volume to a comfortable listening level (see below).

<sup>2</sup> Although using speech-like maskers rather than white noise would be more ecologically valid, given that this is the first use of this novel task and the speech consisted of relatively short isolated words, white noise was used for simplicity.

**Table 1***Description of the Four Experiments in This Study*

Experiment	Stimuli	Noise level	Modality	Secondary task
1	Isolated words, tones	Easy, hard	Audio-only, audiovisual	Tone classification
2	Tones	Easy, hard	Audio-only	Tone classification
3	Isolated words, tones, numbers between 1 and 8	Easy, medium, hard	Audio-only	Tone classification, number classification
4	Isolated words, tones	Nine SNRs	Audio-only, audiovisual	Tone classification

*Note.* SNR = signal-to-noise ratio.

However, the difference was quite small (2.5% across three SNRs), and there was no evidence that the effect of noise type differed across SNRs (see the Supplemental Materials for details).

## Procedure

Prior to beginning the main task in all experiments, participants completed a brief demographic questionnaire asking about their age, biological sex, race, and experience with English. These data are available at <https://osf.io/pqj8h/>. During the instructions phase of each task, participants were reminded that they must wear headphones to complete the task. Participants were then told that on the next screen they would be presented with the loudest noise they would hear during the experiment, and were advised to turn their volume down before hearing the noise, adjust it to a comfortable level as the noise played, and not adjust their volume at any point during the experiment after setting their volume.

After this volume adjustment phase, participants completed a brief tone familiarization phase to ensure that they knew what was meant by “short,” “medium,” and “long” tones. Participants were informed that they would hear each tone twice in a row from short to long and then the short–medium–long sequence three times through (in silence). Participants then completed practice trials followed by the main task.

During each trial, participants were told to press the “J” key as quickly as possible for short tones, the “K” key for medium tones, and the “L” key for long tones. When speech was present (i.e., all experiments except the negative control experiment, Experiment 2), participants were told to respond to the tone first and then type the word they heard in a text box and press the “enter” key to move onto the next trial. In all dual-task experiments, participants were told to complete both tasks to the best of their ability, but to prioritize the speech task, with the assumption being that maintaining successful performance on the speech task in difficult listening conditions draws enough resources away from the secondary task that performance becomes impaired (e.g., Desjardins & Doherty, 2013; Fraser et al., 2010; Hicks & Tharpe, 2002; see Gagné et al., 2017, for a review). The next trial began after a 750-ms interstimulus interval, measured from the time at which the participant pressed the “enter” key following their typed response.

After completing the main task, participants completed a brief questionnaire asking whether they used headphones and whether they adjusted the volume during the experiment. Given that these experiments were conducted remotely and ambient noise could not be controlled, participants who reported using external speakers rather than headphones were excluded from analyses.<sup>3</sup>

## Performance-Based Exclusion Criteria

Unless otherwise noted, participants were excluded if they met any of the exclusion criteria in any condition within a single variable (e.g., SNR) collapsed across the other variable. For example, participants would be excluded for meeting an exclusion criterion in either the easy or hard SNR (collapsed across modality) or in either the audio-only or audiovisual modality (collapsed across SNR). I opted to exclude participants in this manner rather than applying an exclusion criterion to all crossed conditions (i.e., noise-by-modality) to avoid unnecessarily removing participants who performed poorly in just a single subcondition (this decision was preregistered; see <https://osf.io/bqtfy/>). Individual response time trials were excluded if the response time was more than 3 median absolute deviations from the participant’s median response time in that condition.

Participant-level exclusion criteria were as follows: Participants were excluded for having mean speech identification accuracy more than 3 standard deviations below the mean in that condition, mean response times more than 3 standard deviations above or below the mean response time in that condition, or mean tone identification accuracy more than 3 standard deviations below the mean in that condition. For the criteria related to speech identification accuracy, the only exception was if a participant scored 3 standard deviations below the mean in a condition, but their accuracy was still above 90% in that condition. This can occur when the listening conditions are very easy, which produces means near 100% and very small standard deviations. In these cases, participants might have been unnecessarily excluded because the cutoff occurs at an extremely high level of accuracy.

Finally, I excluded participants for failing to complete the speech identification task in any subcondition (i.e., having 0% accuracy in any condition). This exclusion criterion was not preregistered in Experiments 1, 2, or 3 because I did not anticipate that participants would not complete the speech task.<sup>4</sup> However, I opted to exclude these participants before moving forward with any analyses (and before applying other participant-level exclusion criteria) because including them may lead to artificially high accuracy and fast response times as well as increased variability in mean identification accuracy across participants, thereby affecting which participants are excluded for meeting the other performance-based exclusion criteria.

<sup>3</sup> Although some researchers prefer to use a headphone screening task (e.g., Milne et al., 2020; Woods et al., 2017) rather than asking participants about headphone usage at the end of the experiment, the latter approach was used here to reduce the length of the experiment and because participants claiming to use headphones occasionally still fail these screening tasks, which can be frustrating for those individuals.

<sup>4</sup> This exclusion criterion was preregistered in Experiment 4 because preregistration documents for later experiments were submitted after data from the earlier experiments had been analyzed.

## Analysis Plan

Unless otherwise noted in the Results and Discussion sections of the specific experiments, data were analyzed as follows. Response time data were analyzed with linear mixed-effects models assuming a Gaussian distribution. Typed responses were scored in R (R Core Team, 2022) and were hand-checked by the author (blind to condition; see the accompanying R script for details regarding the specific changes made as well as a copy of the raw responses, which were included in addition to the updated responses for transparency). Before scoring responses, I removed extraneous punctuation and converted all responses to lowercase. Responses were scored as correct if they exactly matched the target word, were homophonous with the target word (e.g., “sun” for “son,” “tied” for “tide”), were a common misspelling of the target word (e.g., “theif” for “thief”), or were a single-letter addition, deletion, or substitution of the target word (provided that mismatch was not also a word; e.g., “dfream” for “dream”). Pluralizations were not counted as correct.

Participants and items were included as random effects when appropriate. Given that SNR and modality were manipulated within subjects and within items, I attempted to model by-participant and by-item random slopes for SNR and modality. In cases of non-convergence, I adjusted control parameters (e.g., changed the optimizer, increased the maximum number of iterations) and/or removed correlations among random effects to help enable convergence. However, if the models still would not converge, I removed random slopes that contributed the least to the total variance, and when possible only removed random effects that did not significantly reduce model fit according to likelihood ratio tests. Additional details regarding the particular random effects structures I employed in each experiment are available in the corresponding analysis scripts.

Statistical significance was evaluated by comparing nested models via likelihood ratio tests. Coefficient estimates are provided for effects that were significant according to the relevant likelihood ratio test. To avoid interpreting lower order terms in the presence of interactions, all reported coefficients were derived from models that included the term of interest and all relevant lower order terms, but not higher order terms. For example, a coefficient estimate for a significant two-way interaction comes from a model including that interaction (but no other two-way interactions) as well as all relevant simple effects; in contrast, a coefficient estimate for a significant main effect comes from a model including the effect of interest (e.g., noise level) and all other relevant main effects (e.g., modality), but no interactions. Unless otherwise noted, all models implemented a dummy coding scheme in which the audio-only and easy conditions were coded as 0.

## Experiment 1

The goal of Experiment 1 was to conceptually replicate the finding that audiovisual speech incurs additional dual-task costs relative to audio-only speech in easy but not in difficult listening conditions (Brown & Strand, 2019a) using the novel dual-task paradigm. This experiment also aimed to establish positive control by demonstrating that the novel paradigm is sensitive to changes in SNR such that response times to the secondary task are slower in louder background noise in both audio-only and audiovisual conditions. The preregistration for Experiment 1 is available at <https://osf.io/bqtfy>. Unless otherwise noted in the experiment-specific preregistration documents,

analyses for all subsequent experiments mirror those described in the preregistration for Experiment 1.

## Method

### Transparency and Openness

All data, materials, code, and preregistration documents for this and all experiments in this study can be accessed at <https://osf.io/pqj8h/>. This study conforms to the Transparency and Openness Promotion guidelines outlined by the Open Science Framework.

### Participants

To obtain the final sample size of 88 participants (see preregistration for sample size justification), data from 102 participants were collected from the Washington University in St. Louis Department of Psychological & Brain Sciences subject pool. Before excluding participants for meeting the criteria defined in the preregistration, three participants were excluded for failing to complete the speech identification task in at least one of the four noise-by-modality conditions (i.e., having 0% accuracy). Two participants were excluded for having extreme mean response times in any noise or modality condition, three participants were excluded for having poor speech identification accuracy (suggesting that they were not attending to the task), and seven were excluded for reporting using external speakers rather than headphones to complete the task. These criteria identified 14 unique participants, so the final analysis included data from 88 participants ages 18–24 ( $M_{\text{age}} = 20.0$  years).

### Stimuli

**Words.** A total of 320 words (plus 12 additional words to use in practice trials) were randomly selected from a subset of the SUBTLEX-US database (Brysbaert et al., 2012), excluding articles and conjunctions, uncommon words (log-frequencies less than three), and long words (more than two syllables or five phonemes). Audiovisual stimuli were recorded by a female American English speaker without a discernible regional accent. Auditory stimuli were recorded at 16-bit, 44100 Hz using a Shure KSM-32 microphone with a plosive screen, and visual stimuli were recorded with a Panasonic AG-AC90 camera. These long audiovisual recordings were split into isolated word files with approximately 330 ms of silence before the onset of the word (though the precise onset of the speech relative to the onset of the file varied). Audio-only stimuli were created by adding a black screen to the audiovisual files to ensure that the file types were the same across conditions in case video and audio files load at different rates in the experiment presentation software.

**Tones and Noise.** The tone and noise files were mixed first (because these stimuli would be used in Experiment 2), and then these files were mixed with the speech files to generate audio-only and audiovisual versions of all stimuli. In the unmixed speech files, the words occurred approximately 330 ms after the onset of the file. To ensure that the speech did not begin too abruptly in the mixed files, the noise began 250 ms before the onset of the speech file (and continued throughout it); in other words, the speech began approximately 580 ms after the onset of the mixed stimulus files (250 ms of noise before the onset of the speech file + an additional ~330 ms of



noise before the speech began). The tones occurred at one of four points relative to the onset of the mixed stimulus file: 580, 770, 960, and 1,150 ms (i.e., 190-ms intervals). Thus, the tone occurred at approximately the following time points relative to the onset of the speech: 0, 190, 380, and 570 ms. This timing ensured that the tone and speech typically overlapped and that the decision processes for responding to the two tasks coincided. Particular words were yoked to tones such that for a given word, the tone was always the same length and occurred at the same time point regardless of modality or noise level. Precise details regarding stimulus creation—including notes, ffmpeg code, and additional stimulus details—can be found in the “Creating Stimuli Notes” document at <https://osf.io/pqj8h/>. The easy noise was leveled to  $-36$  dB RMS, and the hard noise was leveled to  $-12$  dB RMS to generate easy (approximately  $+12$  dB SNR) and hard (approximately  $-12$  dB SNR) levels of background noise. The speech was set to  $-24$  dB RMS unless otherwise noted.

### Procedure

Participants completed 320 trials: 80 per condition (audio-only vs. audiovisual; easy vs. hard). Modality and noise level were blocked, and the order of the blocks was counterbalanced across participants. Within each block, the words were presented in a randomized order. All words appeared in all conditions across participants, but each participant heard each word only once. Participants were familiarized with the tones as described in the General Method and Analysis Plan section and then completed 12 practice trials (three per condition). The experiment took approximately 45 min to complete.

### Results and Discussion

A total of 2,784 trials (8.5% of the data) with response times more than 3 median absolute deviations from the median response time for a particular participant in a particular condition were removed from further analysis (note that this preregistered criterion was applied before any other exclusion criteria were applied; see preregistration). In the final data set of 88 participants, mean accuracy at classifying the three tones was 52.6% across conditions—poorer than expected but well above the chance level of 33.3% (see Table 2). These incorrect trials were removed prior to analysis for this and all following experiments.<sup>5</sup> The final analysis included 13,791 response times from 88 participants.

Following the analysis plan outlined in the preregistration, I first assessed the effect of background noise in the audio-only and audiovisual conditions separately to ensure that the task was sensitive to changes in SNR—which have repeatedly been shown to affect response times to secondary tasks in the listening effort literature—in both modalities (i.e., establish positive control). Likelihood ratio tests comparing nested models differing only in the presence of a fixed effect for SNR (easy vs. hard) indicated that SNR significantly affected response times to the tone classification task in both the audio-only ( $\chi^2_1 = 22.41$ ,  $p < .001$ ) and audiovisual conditions ( $\chi^2_1 = 7.72$ ,  $p = .005$ ). More difficult levels of background noise slowed response times by an estimated 199 ms in the audio-only condition ( $B = 199.48$ ,  $SE = 39.68$ ,  $t = 5.03$ ,  $p < .001$ ; Cohen’s  $d = 0.33$ ) and 112 ms in the audiovisual condition ( $B = 112.46$ ,  $SE = 39.68$ ,  $t = 2.83$ ,  $p = .006$ ; Cohen’s  $d = 0.22$ ). These results are consistent with previous work demonstrating that

identifying speech in difficult levels of background noise slows response times to unrelated tasks (e.g., Brown & Strand, 2019a; Sarampalis et al., 2009). Crucially, these findings establish positive control by demonstrating that the novel dual-task paradigm is sensitive to changes in the level of the background noise when speech is present; indeed, most participants had slower mean response times in the hard than the easy noise level (collapsed across modalities; see Figure 2).

The analyses above indicated that the novel dual-task paradigm can reliably detect changes in the dual-task costs associated with processing speech in easy versus difficult noise levels, regardless of whether the listener can see the talker’s face. The next set of analyses conceptually replicated our previous finding that the effect of modality on secondary task response times differs depending on the level of the background noise (Brown & Strand, 2019a). A likelihood ratio test comparing nested models differing only in the presence of the interaction term indicated that modality differentially affected response times to the secondary task in the two noise levels ( $\chi^2_1 = 30.89$ ,  $p < .001$ ). Examination of the summary output for the full model (i.e., the model including the interaction term) indicated that the effect of modality was significantly more negative in the hard than the easy condition, meaning that participants responded more quickly in the audiovisual condition than the audio-only condition, particularly in hard background noise ( $B = -82.55$ ,  $SE = 14.84$ ,  $t = -5.56$ ,  $p < .001$ ). Given that the interaction was significant, I also assessed the effect of modality in each noise level separately. These analyses indicated that the modality effect was significant in the hard noise level ( $\chi^2_1 = 4.06$ ,  $p = .04$ ) such that response times were an estimated 75 ms faster in the audiovisual relative to the audio-only condition ( $B = -75.01$ ,  $SE = 36.96$ ,  $t = -2.03$ ,  $p = .046$ ), but modality did not significantly affect response times in the easy condition ( $\chi^2_1 = 0.04$ ,  $p = .85$ ). These data are shown in Figure 3 and Table 2 (see the Supplemental Materials for alternative visualizations).

Finally, to determine whether response times were affected by noise level and modality overall, I compared a model including both variables (but no interaction term) to nested models lacking either noise level or modality as a fixed effect. These analyses indicated a main effect of noise level ( $\chi^2_1 = 22.60$ ,  $p < .001$ ) such that participants responded an estimated 160 ms more slowly when speech was presented in hard relative to easy background noise levels, controlling for modality ( $B = 160.20$ ,  $SE = 31.71$ ,  $t = 5.05$ ,  $p < .001$ ). The main effect of modality was not significant ( $\chi^2_1 = 2.06$ ,  $p = .15$ ).

Taken together, the results of Experiment 1 are consistent with previous work showing that seeing the talker differentially affects dual-task costs depending on the level of the background noise. In our previous study comparing dual-task response times for audio-only and audiovisual speech in easy and hard levels of background noise, we found that audiovisual speech led to *slower* response times than audio-only speech, but only in the easy SNR (Brown & Strand, 2019a). Here, I found that audiovisual speech led to *faster* response times than audio-only speech, but only in the hard SNR. Although these results may seem incompatible, they are perfectly consistent with the theoretical account (i.e., multiple resource theory) outlined

<sup>5</sup> The decision to exclude trials that were incorrectly classified as short, medium, or long from analyses was made prior to data collection, following the conventions of previous research using a similar paradigm (Brown & Strand, 2019a).



**Table 2**

*Mean Tone Identification Accuracy, Word Identification Accuracy, and Response Time to the Tone Task in Each of the Four Conditions in Experiment 1*

Task condition	Accuracy on tone task (%)	Word identification accuracy (%)	Response time to tone task (ms)
AO, easy	54.3	94.0	1,181
AO, hard	50.8	60.5	1,372
AV, easy	54.8	95.1	1,183
AV, hard	50.6	79.4	1,309

*Note.* AO = audio-only; AV = audiovisual.

above: Audiovisual speech may incur a small processing cost in easy listening conditions—perhaps as a result of distraction, audiovisual integration costs, monitoring two channels, and so forth—but these costs are offset by the benefit of reduced lexical competition in difficult listening conditions. Explanations for the discrepancies between the studies are explored in detail in the General Discussion section.

## Experiment 2

The goal of Experiment 2 was to establish negative control by demonstrating that the tone classification task is *not* sensitive to changes in noise level in the absence of speech. If response times are similarly slowed by the presence of louder background noise regardless of whether speech is present or absent, this complicates the interpretation that the effects observed in Experiment 1 are attributable to increased cognitive demands associated with identifying speech in noise. If, however, response times are slower in the louder background noise level when speech is present (Experiment 1) but *not* when speech is absent (Experiment 2), this suggests that the effects observed in Experiment 1 cannot be attributable to either noise-induced cognitive interference (see, e.g., Brown & Strand, 2019b; Colle & Welsh, 1976) or inaudibility of the tone in louder

background noise (which would be unexpected given the limited spectral overlap between the tone and noise, but not impossible). The preregistration for Experiment 2 is available at <https://osf.io/xbme6>.

## Method

### Participants

Seventy-five individuals from the Washington University in St. Louis Department of Psychological & Brain Sciences subject pool participated in this study prior to exclusion. One participant was excluded for responding before the tone on every trial in one condition, indicating that they were not attending to the task. This criterion was not preregistered because I did not anticipate that any participants would ignore the instructions and respond in this manner. Three participants were excluded for having extreme mean response times in either condition, one was excluded for having poor tone identification accuracy, and five were excluded for reporting using external speakers rather than headphones to complete the task. These preregistered criteria identified seven unique participants (in addition to the one who failed to attend to the task), so the final analysis included data from 67 participants. Note that I preregistered a sample size of 65 but indicated that I would continue collecting data as long as it was feasible to do so, and I would not conduct any statistical analyses until data collection was complete. I therefore opted to use the data from all 67 participants to increase statistical power (ages 18–35;  $M_{\text{age}} = 20.2$  years).

### Stimuli

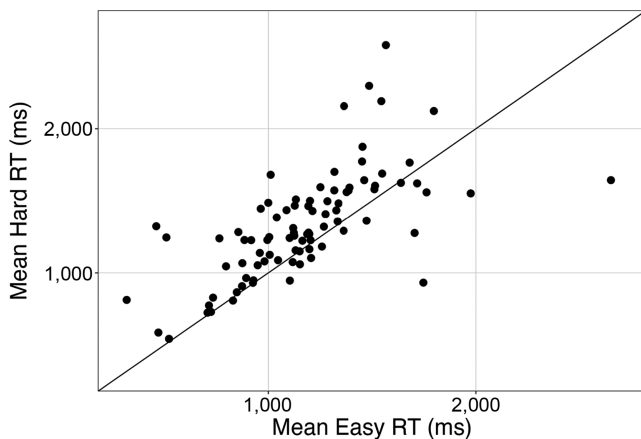
This experiment used the same background noise and tones as those in Experiment 1, but the tones in noise were presented without speech.

### Procedure

The procedures for this task were identical to those in Experiment 1, but the tone and noise were not mixed with speech, so modality was not relevant in this experiment. Participants classified 240 tones (120 per noise level) as short, medium, or long as quickly as possible. Noise level was blocked, and the order of the blocks was counterbalanced across participants. Prior to completing the main task, participants were familiarized with the tones and then completed 12 practice trials. The experiment took approximately 25 min to complete.

**Figure 2**

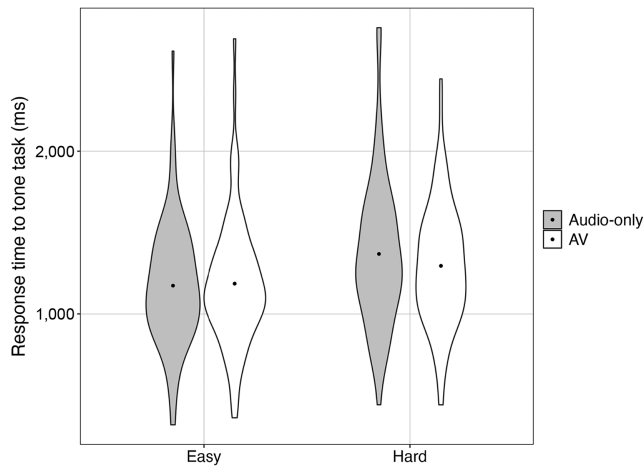
*Scatterplot Showing Average RTs to Classify Tones as Short, Medium, or Long for Each Participant in the Easy and Hard Noise Levels (Collapsed Across Modality) in Experiment 1*



*Note.* The fact that most points are above the line  $y = x$  demonstrates that participants' RTs were systematically slower in the harder noise level. RT = response time.

**Figure 3**

*By-Participant Response Times to the Secondary Tone Classification Task for the Easy (Left) and Hard (Right) Signal-to-Noise Ratios in Experiment 1*



*Note.* The dot represents the mean response time in each condition, and the shape of each plot depicts the distribution of responses across participants. AV = audiovisual.

## Results and Discussion

A total of 1,562 trials (8.7% of the data) with response times more than 3 median absolute deviations from the median response time for a particular participant in a particular condition were removed from further analysis. Mean accuracy at classifying the three tones was 67.3% in the easy condition and 63.9% in the hard condition (65.6% across conditions). Incorrect trials were removed prior to analysis. The final analysis included 9,948 response times.

To assess whether the level of the background noise affected response times to the tone task when speech was not present, I first built a full model including noise level (easy vs. hard) as a fixed effect, random intercepts for participants, and by-participant random slopes for noise level. Given that this experiment included only three unique items, I did not attempt to model item-level random effects. This model was compared to a reduced model lacking the fixed effect for noise level via a likelihood ratio test, which indicated that the effect of noise level was not significant ( $\chi^2_1 = 0.01, p = .92$ ; see Figure 4). Indeed, the mean response time in the easy SNR (803 ms) was nearly identical to that in the hard SNR (798 ms). Thus, consistent with previous work using a different dual-task paradigm (Brown & Strand, 2019b), I did not find evidence for a noise-related difference in response times to the tone classification task when speech was not present. Taken together with the results of Experiment 1, these findings suggest that the differences in response times between the easy and hard SNRs observed in Experiment 1 are not attributable to noise-induced cognitive interference or to reduced audibility of the tones in the hard noise level. Instead, it appears that the additional recruitment of cognitive resources that is required to identify spoken words in high levels of background noise leaves fewer resources available to complete simultaneous cognitive tasks, which in turn leads to slower response times to the secondary task.

## Experiment 3

Experiment 3 evaluated the sensitivity of the novel dual-task paradigm to changes in SNR (using three SNRs: easy, moderate, and hard). Here, “sensitivity” refers to the extent to which changes in SNR produce changes in response times to the secondary task, where larger response time differences for a fixed change in SNR indicate greater measurement sensitivity. This experiment also evaluated the novel task’s convergent validity relative to a task that is commonly used in the listening effort literature (see Brown & Strand, 2019b; Picou & Ricketts, 2014; Sarampalis et al., 2009). This previous task involves classifying numbers as even or odd; however, given that the numbers are presented visually on the screen, this task is incompatible with audiovisual speech research because the screen can only be occupied by one task at a time. Thus, to enable comparison between the two tasks, Experiment 3 was conducted in the auditory modality alone. The preregistration documents for Experiment 3, which include minor additions that were added before data were collected, are available at <https://osf.io/2jmqz> and <https://osf.io/uvnax>.

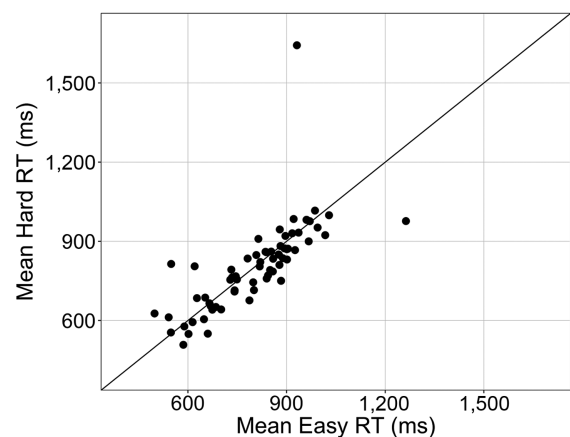
## Method

### Participants

To reach the preregistered sample size of 100 participants, data were collected from 117 participants via the Washington University in St. Louis Department of Psychological & Brain Sciences subject pool and Prolific. I excluded participants who did not provide complete data sets for both tasks. For the tone task, two participants were excluded for failing to complete the speech identification task in one of the three SNRs (no participants met the other exclusion criteria). For the number task, two participants were excluded for failing to complete the speech identification task in one of the three SNRs,

**Figure 4**

*Scatterplot Showing Average RTs to Classify Tones as Short, Medium, or Long for Each Participant in the Easy and Hard Noise Levels in Experiment 2*



*Note.* The fact that the points are centered around the line  $y = x$  demonstrates that participants’ RTs were not systematically affected by the level of the background noise when speech was not present. RT = response time.

four participants were excluded for having poor speech identification accuracy, six participants were excluded for having extreme mean response times, and one participant was excluded for having poor accuracy at classifying the numbers as even or odd (these criteria identified 11 unique participants). Across both tasks, eight participants were excluded for using external speakers rather than headphones. The final analysis included data from 100 participants ages 18–30 ( $M_{\text{age}} = 24.1$  years).

### Stimuli

Target words consisted of a subset of 240 audio-only words from a previous study (Strand, Brown, & Barbour, 2020) that were no more than two syllables or five phonemes in length and excluded conjunctions, articles, and uncommon words (log-frequencies less than three). Half of these words (120) were randomly assigned to be used in the tone task, and the other half were used in the number task, resulting in 40 words per noise level in each task. The background noise was the same for both tasks and consisted of the same notched noise used in both previous experiments presented at three SNRs: easy (12 dB), moderate (0 dB), and hard (−12 dB).

The tones in Experiment 3 were identical to those used in the previous experiments in this study and occurred at the same points relative to the onset of the stimulus file and the onset of the speech. Stimuli for the number task consisted of images of two side-by-side boxes (measuring approximately 5 cm per side, but this changed depending on the size of the participant's monitor), one of which contained a digit between 1 and 8, as well as an image of the same two boxes without any digits. Tones were not presented in the odd/even classification task.

### Procedure

Participants completed 320 trials: 40 trials for each of the three noise levels in each task, plus 40 single-task trials in each task (i.e., the tone and odd/even classification tasks without speech). The tasks were blocked, and the order of the blocks was counterbalanced across participants such that approximately half of the participants completed the tone task first and half completed the number task first (but all participants completed both tasks). Within each task, SNR was blocked and the order of the SNRs was counterbalanced across participants, but the single-task trials always occurred last within each tone or number block. Within any given block, the order in which the trials were presented was randomized. Across participants, all words appeared in all SNRs (note that different words were used for the two tasks), but each participant heard each word only once. The experiment took approximately 40 min to complete.

The procedure for the tone task was identical to the procedure previously described, but after completing dual-task trials at each SNR, participants completed the tone task in isolation (i.e., without words) in background noise at a level corresponding to the moderate SNR (0 dB). Prior to beginning the task, participants were familiarized with the tones and then completed 12 practice trials.

The number task followed the procedures used in previous work (e.g., Brown et al., 2020; Brown & Strand, 2019b; Sarampalis et al., 2009). In this task, participants made speeded judgments indicating whether visually presented integers between 1 and 8 were even or odd while simultaneously identifying words in the notched background noise. If the number was even, participants were instructed

to press a key corresponding to an arrow pointing *toward* the box with that number, and if the number was odd, participants were instructed to press a key corresponding to an arrow pointing *away* from the box with that number. The “F” key represented a left-facing arrow, and the “J” key represented a right-facing arrow. For example, if the number 4 appeared on the left, participants responded by pressing the “F” (left) key, but if the number 5 appeared on the left, they responded by pressing the “J” (right) key. The appearance of the number roughly coincided with the presentation of the words, but the numbers appeared after a variable delay for each trial ranging from 750 to 1,150 ms after trial onset (in 100-ms intervals). Participants were instructed to respond to the number task first and then type the word they heard in a text box. After completing the three blocks of dual-task trials, participants completed the number task without also hearing words. These single-task trials occurred in background noise at a level corresponding to the moderate SNR (0 dB). The dependent variable of interest was response time to the number task. Participants completed 12 practice trials prior to beginning the task.

## Results and Discussion

### Tone Task

A total of 1,095 trials (5.8% of the data) with extreme response times were removed prior to analysis. Mean accuracy at classifying the three tones was 59.1% across conditions (see Table 3). After removing incorrect trials, the final analysis included 6,281 dual-task and 2,637 single-task response times from 100 participants.

To confirm that the tone task was sensitive to changes in SNR, I compared a model with a fixed effect for noise level to a model lacking that fixed effect using only data from dual-task trials from the tone task. This likelihood ratio test indicated that the effect of noise level was significant ( $\chi^2_2 = 43.58, p < .001$ ). Indeed, relative to the hardest noise level, response times were on average an estimated 173 ms faster in the moderate noise level ( $B = -172.75, SE = 27.04, t = -6.39, p < .001$ ) and 197 ms faster in the easy noise level ( $B = -197.36, SE = 27.69, t = -7.13, p < .001$ ).<sup>6</sup> Finally, I tested whether participants responded more quickly in single-task relative to dual-task conditions by subsetting the dual-task data to only include the 0-dB SNR condition (because this was the noise level that was presented during single-task trials). A likelihood ratio test indicated that response times differed significantly between dual-task and single-task trials ( $\chi^2_1 = 65.19, p < .001$ ), and examination of the summary output for the full model indicated that participants responded an estimated 430 ms more quickly in single-task relative to dual-task trials ( $B = -430.15, SE = 44.96, t = -9.57, p < .001$ ).

### Number Task

A total of 1,422 trials (7.6% of the data) with extreme response times were removed prior to analysis. Mean accuracy at classifying the numbers as even or odd was 83.0% across conditions (see Table 3). After removing incorrect trials, the final analysis included 11,093 dual-task and 3,718 single-task trial response times from the same 100 participants who completed the tone task.

<sup>6</sup> The SNR effect also emerged when noise was coded numerically rather than categorically. See the Supplemental Materials.

**Table 3**

*Mean Tone and Number Classification Accuracy, Word Identification Accuracy in Each Task, and RT to the Secondary Task for Both Tasks in Each of the Four Conditions in Experiment 3*

Task condition	Tone task accuracy (%)	Number task accuracy (%)	Word accuracy, tone task (%)	Word accuracy, number task (%)	Tone task RT (ms)	Number task RT (ms)
Easy (12 dB)	58.1	81.6	95.2	95.1	1,149	899
Moderate (0 dB)	57.4	82.4	88.5	89.3	1,179	906
Hard (-12 dB)	50.6	83.6	53.7	63.3	1,358	1,062
Single-task (0 dB)	70.2	84.2			778	687

*Note.* Words were not presented in the single-task trials, so these cells in the table are empty. RT = response time.

As above, a likelihood ratio test confirmed that background noise level significantly affected dual-task response times ( $\chi^2_2 = 35.51, p < .001$ ). Relative to the hardest noise level, response times were on average an estimated 156 ms faster in the moderate noise level ( $B = -155.92, SE = 24.62, t = -6.33, p < .001$ ) and 162 ms faster in the easy noise level ( $B = -161.90, SE = 28.87, t = -5.61, p < .001$ ; Footnote 6). Finally, a likelihood ratio test indicated that response times differed significantly between dual-task (0-dB SNR only) and single-task trials ( $\chi^2_1 = 45.37, p < .001$ ) such that participants responded an estimated 217 ms more quickly in single-task relative to dual-task trials ( $B = -216.90, SE = 28.77, t = -7.54, p < .001$ ).

### Comparing the Tone and Number Tasks

**Sensitivity and Convergent Validity.** Given that there are no clear conventions regarding the best way to assess sensitivity and convergent validity for dual-task paradigms across levels of background noise, I employed multiple methods in this experiment (see preregistration document). First, I used the dual-task data to calculate Pearson correlations between by-participant mean response times for the two tasks at each noise level. All three of these correlations were significant and increased in magnitude as the level of the background noise increased (see the first column of Table 4). Interestingly, the correlation between single-task mean response times for the two tasks was small and nonsignificant ( $r = .11, p = .26$ ). This is somewhat surprising given that both the tone and number tasks rely on processing speed, so it might be expected that participants who tend to respond quickly on one task also respond quickly on the other. Instead, the correlation between mean response times for the two tasks only emerged when the tasks were completed while simultaneously identifying speech in noise, and the relationship became stronger as the speech task became more difficult. Explanations for this pattern of results are explored in the General Discussion section.

Next, I calculated *proportional dual-task costs* for each task at each noise level, which indicate the extent to which participants are negatively affected by dual-tasking at each noise level relative to their single-task performance (Gosselin & Gagné, 2011a, 2011b). For each task and each of the three SNRs, I calculated by-participant proportional dual-task costs using the following equation: (dual – single)/single. Here, “dual” and “single” refer to the mean dual- or single-task response time in a particular condition for a particular participant. Proportional dual-task costs are assumed to reveal the degree of impairment as a result of dual-tasking at each SNR, normalized by single-task performance to control for any hardware differences that might affect raw response times, as well as control for individual differences in processing speed. At each SNR, the mean proportional dual-task cost was larger for the tone task than the number task, suggesting that the tone task is more sensitive to changes in SNR than the number task (see the first two columns of Table 5). Additionally, for the most part, the mean proportional dual-task cost within each task increased as the SNR became more difficult (though the mean cost was similar for the number task in the easy and moderate noise levels), indicating that both tasks were sensitive to changes in SNR. However, correlation analyses revealed that none of the correlations involving proportional dual-task costs were significant (see the second column of Table 4), suggesting that there is no evidence that the extent to which a participant is negatively affected by dual-tasking in one task is related to the extent to which they are negatively affected by dual-tasking in another, according to the proportional dual-task cost metric.

Next, I calculated proportional dual-task costs relative not to single-task performance, but rather to performance in the easier adjacent noise level (see the last two columns of Table 5). That is, I calculated the cost of moving from the easy to the moderate noise level and the cost of moving from the moderate to the hard noise level for each task. Although proportional dual-task costs increased as the level of the background noise increased for both

**Table 4**

*Pearson Correlations Between RTs to the Tone and Number Tasks on Three Scales: Raw, pDTC Relative to Single-Task Performance, and pDTC Relative to Performance in the Easier Adjacent Noise Level*

Task condition	Raw RT	pDTC (single-task)	pDTC (adjacent)	RT difference (single-task)	RT difference (adjacent)
Easy (12 dB)	<b>0.36</b> ( $p < .001$ )	-0.09 ( $p = .36$ )		0.15 ( $p = .14$ )	
Moderate (0 dB)	<b>0.47</b> ( $p < .001$ )	-0.10 ( $p = .33$ )	0.17 ( $p = .08$ )	<b>0.41</b> ( $p < .001$ )	0.20 ( $p = .0504$ )
Hard (-12 dB)	<b>0.56</b> ( $p < .001$ )	-0.05 ( $p = .65$ )	0.05 ( $p = .61$ )	<b>0.39</b> ( $p < .001$ )	0.11 ( $p = .28$ )

*Note.* Adjacent pDTCs cannot be calculated for the easiest noise level, so these cells in the table are empty. Values in bold indicate significant correlations. RT = response time; pDTC = proportional dual-task cost.



**Table 5***Mean pDTC for the Tone and Number Tasks, Relative to Both Single-Task Performance and Performance in the Easier Adjacent Noise Level*

Task condition	Tone task mean pDTC (relative to single-task)	Number task mean pDTC (relative to single-task)	Tone task mean pDTC (relative to adjacent SNR)	Number task mean pDTC (relative to adjacent SNR)
Easy (12 dB)	0.37	0.31		
Moderate (0 dB)	0.40	0.30	0.03	0.01
Hard (−12 dB)	0.61	0.53	0.17	0.18

*Note.* The pDTCs relative to the adjacent noise level cannot be calculated for the easiest noise level. pDTC = proportional dual-task cost; SNR = signal-to-noise ratio.

tasks, correlation analyses failed to find evidence that participants are similarly affected by changes in the level of the background noise across the two tasks (see the third column of Table 4).

In an exploratory analysis, I calculated the difference in mean response times between single- and dual-task trials at each noise level for each task (this value is similar to proportional dual-task costs but is on the raw response time scale rather than the normalized scale) and then correlated these raw response time differences across tasks at each noise level. With this outcome, the correlations were significant in the moderate and hard noise levels, but not in the easy noise level (see the fourth column of Table 4). I also conducted a similar set of exploratory analyses using differences in dual-task response times between adjacent noise levels (rather than between dual- and single-task trials, as in the proportional dual-task cost analysis); none of these correlations were significant (see the last column of Table 4).

In the final set of preregistered analyses, I calculated Cohen's  $d$  values indicating the magnitude of the effect of dual-tasking at each noise level relative to single-task response times and relative to response times in the adjacent noise level (note that the latter values reflect the magnitude of the effect of SNR rather than dual-tasking; Table 6). Consistent with the proportional dual-task cost results above (Table 5), these analyses indicated that effect sizes increased as the level of the background noise increased from easy to hard for both tasks (though again, this was not the case for the easy and moderate levels of background noise in the number task), and all effect sizes were larger in the tone task than the number task.

**Mixed-Effects Regression Analysis.** Another method of assessing the sensitivity of the two tasks is to move beyond participant-level analyses that rely on means and correlations, and instead use a mixed-effects regression framework. Specifically, the next set of analyses evaluated whether the two paradigms are similarly affected by changes in SNR by testing the interaction between noise level (easy, moderate, hard) and task (tone vs. number) using only dual-task trials from both tasks. A likelihood ratio test indicated that the interaction was significant ( $\chi^2 = 11.19$ ,  $p = .004$ ) such that the magnitude of the change in response times from the hard to the easy noise level differed significantly between the two tasks ( $B = -48.09$ ,  $SE = 14.38$ ,  $t = -3.34$ ,  $p < .001$ ), but the magnitude of the change from the hard to the moderate noise level did not ( $B = -23.39$ ,  $SE = 14.40$ ,  $t = -1.62$ ,  $p = .10$ ). Overall, the interaction indicated that although participants tended to respond more quickly in the easy relative to the hard noise level in both tasks, this effect was stronger in the tone task. These results are consistent with the participant-level analyses and

suggest that the tone task is more sensitive to changes in SNR than the number task.<sup>7</sup>

Taken together, the results of Experiment 3 provide converging evidence that the tone task is more sensitive to changes in noise level than the number task. This experiment also provided some convergent validity evidence for the tone task: Response times to both tasks increased as the level of the background noise increased, significant correlations were observed between mean response times on the two tasks at all three noise levels, and the magnitudes of the correlations became much stronger as the background noise level increased. Consistent with the correlation analyses of the raw (i.e., un-normalized) response time data, the response time curves for the two tasks across noise levels looked nearly identical (see Figure 5). However, in contrast to the analyses of the raw response time data, when response times were normalized (either relative to single-task response times or response times in the easier adjacent noise level), all of these correlations disappeared. Explanations for the discrepancies among the results using raw response times, difference scores, and proportional dual-task costs are discussed in detail in the General Discussion section. Finally, both tasks were more sensitive to changes in SNR between the moderate and hard conditions than between the easy and moderate conditions (see Figure 5), likely because differences in speech intelligibility were more pronounced between the moderate and hard noise levels.

## Experiment 4

The goal of Experiment 4 was to generate performance curves for both word identification accuracies and response times to the tone classification task in audio-only and audiovisual conditions. The results of this experiment will provide other researchers with valuable information about the SNRs at which the novel task is likely to be most sensitive. It will also allow researchers to compare performance on the tone task across audio-only and audiovisual conditions across a wide range of SNRs, which will help address questions regarding dual-task costs for audio-only and audiovisual speech when either SNR or speech intelligibility is matched across

<sup>7</sup> An exploratory analysis mirroring the mixed-effects analysis described above indicated that including a covariate representing each participant's mean single-task response time collapsed across the two tasks did not affect the results. Indeed, the interaction was significant in this analysis as well and the estimate for the interaction term was identical within three digits to the estimate in the previous analysis ( $\chi^2 = 11.19$ ,  $p = .004$ ). An additional exploratory analysis revealed that these effects were still significant when SNR was coded numerically rather than categorically (see the Supplemental Materials).

**Table 6**

*Cohen's  $d$  Values for Response Times in Each Task at Each Noise Level Relative to Response Times in Single-Task Trials, as Well as Cohen's  $d$  Values Relative to the Easier Adjacent Noise Level*

Task condition	Tone task Cohen's $d$ (relative to single-task)	Number task Cohen's $d$ (relative to single-task)	Tone task Cohen's $d$ (relative to adjacent SNR)	Number task Cohen's $d$ (relative to adjacent SNR)
Easy (12 dB)	0.93	0.57		
Moderate (0 dB)	0.95	0.55	0.06	0.01
Hard (-12 dB)	1.24	0.78	0.32	0.27

*Note.* Cohen's  $d$  values relative to the adjacent noise level cannot be calculated for the easiest noise level. SNR = signal-to-noise ratio.

modalities. The preregistration for Experiment 4 is available at <https://osf.io/dn7ek>.

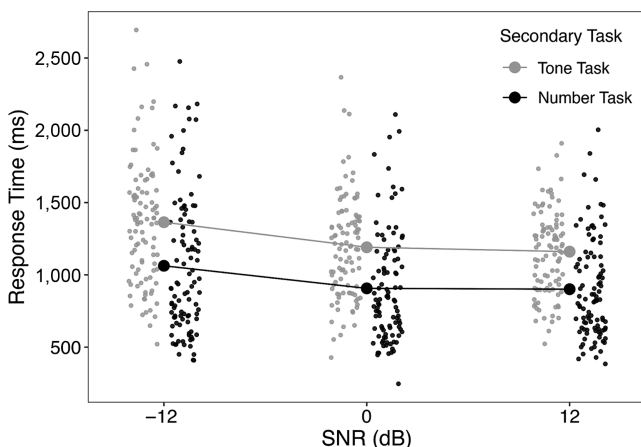
## Method

### Participants

To reach a preregistered sample size of 200 participants (100 per modality), I collected data from 210 participants (104 audio-only and 106 audiovisual) via the Washington University in St. Louis Department of Psychological & Brain Sciences subject pool and Prolific. In the audio-only condition, one participant was excluded for failing to complete the speech identification task, two participants were excluded for having poor speech identification accuracy, and two participants were excluded for using external speakers rather than headphones. These exclusion criteria identified four unique participants, so the audio-only analysis included data from 100 participants ages 18–30 ( $M_{\text{age}} = 22.9$  years). In the audiovisual condition, one participant was excluded for having poor speech identification accuracy, two participants were excluded for having extreme mean response times, and three participants were excluded for using external speakers rather than headphones. These criteria identified six unique participants, so the audiovisual analysis included data from 100 participants ages 18–30 ( $M_{\text{age}} = 23.3$  years).

**Figure 5**

*Mean Response Times (Large Circles) Across the Three Levels of Background Noise in Each Task (Tone vs. Number) in Experiment 3*



*Note.* Smaller circles represent the by-participant mean response times in each condition. SNR = signal-to-noise ratio.

### Stimuli

Stimuli in Experiment 4 came from the same set of recordings as those in Experiment 1, but here I randomly selected 270 words (plus 12 additional words to use in practice trials) that were distinct from those used in Experiment 1. The tones in Experiment 4 were identical to those used in previous experiments in this study and occurred at the same points relative to the onset of the stimulus file and the onset of the speech. The tones were mixed with the same notched background noise as in the previous experiments, but the background noise was presented at nine levels ranging from approximately -12 to +12 dB SNR in 3 dB increments.

### Procedure

Participants completed 270 trials of the dual-task paradigm, 30 per noise level. Given that the goal of the experiment was to generate performance curves for audio-only and audiovisual speech across noise levels—not to directly compare performance across modalities—modality was manipulated between subjects. Thus, participants completed 30 trials in each of the nine noise levels in either the audio-only or audiovisual conditions. Noise level was blocked, the order of the blocks was counterbalanced across participants, and trial order was randomized within each block. Participants were familiarized with the tones and then completed 12 practice trials (at least one per noise level). The experiment lasted approximately 30 min.

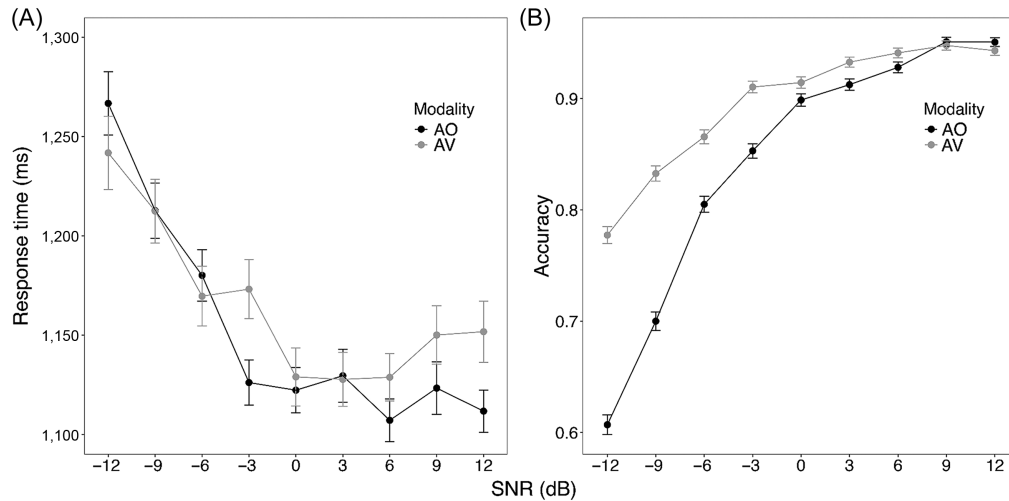
### Results and Discussion

A total of 1,615 trials (5.8% of the data) with extreme response times were removed from the audio-only data, and 1,809 trials (6.3% of the data) with extreme response times were removed from the audiovisual data prior to analysis. Mean tone identification accuracy was 55.0% in the audio-only condition and 53.4% in the audiovisual condition. Incorrect trials were removed prior to analysis. The final analyses included 13,990 audio-only response times and 13,519 audiovisual response times from 100 audio-only participants and 100 unique audiovisual participants.

Performance curves showing mean response times and accuracies at each of the nine noise levels for audio-only and audiovisual words are shown in Figure 6 (see the Supplemental Materials for alternative visualizations). These plots demonstrate that for both audio-only and audiovisual words, as the SNR becomes easier, response times to the secondary task become faster and intelligibility improves. However, in the response time plot (Figure 6A), there appears to be a “knee point” such that after a particular SNR, response times are

**Figure 6**

Mean AO and AV Response Times (Panel A) and Accuracies (Panel B) Across the Nine SNRs Included in Experiment 4



Note. Error bars represent 1 standard error of the mean. AO = audio-only; AV = audiovisual; SNR = signal-to-noise ratio.

largely unaffected by further changes in SNR. For example, in the audio-only condition, each 3 dB increase in SNR is associated with a relatively large drop in response times up until approximately -3 dB SNR (which corresponds to ~85% accuracy), at which point the effect of SNR on response times asymptotes such that the same 3 dB increase in SNR has little effect on response times to the tone task (and produces small but nonnegligible increases in intelligibility). This trend is similar for the audiovisual data, but the asymptote occurs at approximately 0 dB SNR (or ~91% accuracy). Thus, if researchers are interested in detecting changes in response times using this secondary tone classification task and an audio-only speech task, the tone task will be most sensitive if they choose SNRs that result in speech identification accuracies below approximately 85%.

Examination of Figure 6A reveals that participants tended to respond to the tone classification task more quickly in audiovisual than audio-only conditions in the most difficult SNR, but responded comparably or more slowly in audiovisual than audio-only conditions in the easiest SNRs. To assess whether this visual trend and the overall effect of SNR were statistically significant, I conducted several exploratory analyses. The first aimed to evaluate whether response times were significantly affected by SNR by comparing a model with a fixed effect for noise level to a model lacking that fixed effect in each modality separately. Response times differed by noise level in both the audio-only ( $\chi^2_8 = 26.38, p < .001$ ) and audiovisual modalities ( $\chi^2_8 = 16.85, p = .03$ ). In both cases, more difficult levels of background noise led to systematically slower response times to the tone task.<sup>8</sup>

These exploratory analyses replicate previous work demonstrating that response times to secondary tasks tend to be slower when speech is presented in more difficult levels of background noise, regardless of whether the speech is presented in audio-only or audiovisual conditions (see, e.g., Experiment 1 of the present study). In the next set of analyses, I combined the audio-only and audiovisual data and tested the main effects of background noise and modality.

These analyses mirror those in Experiment 1, but in this experiment, modality was manipulated between subjects rather than within subjects, and this experiment included nine rather than two SNRs. For the following analyses, noise level and modality were dummy coded with the hardest noise level and the audio-only condition as the reference levels. Consistent with previous work and my findings from Experiment 1, the effect of SNR was significant in the combined audio-only and audiovisual data, controlling for modality ( $\chi^2_8 = 297.79, p < .001$ ). The main effect of modality was not significant ( $\chi^2_1 = 0.00, p = .99$ ).

Next, I tested the interaction between noise level and modality via a likelihood ratio test. The interaction was significant ( $\chi^2_8 = 19.47, p = .01$ ), suggesting that the effect of modality differed across noise levels.<sup>9</sup> Examination of the summary output indicated that the interaction term comparing the modality effect in the easiest (12 dB SNR) relative to the hardest (-12 dB SNR) noise level was significant and in the same direction as in Experiment 1 ( $B = 62.08, SE = 21.84, t = 2.84, p = .004$ ; these statistics correspond to the coefficient for the interaction between modality and the easiest of the nine SNRs). Consistent with the results of Experiment 1, seeing the talker tended to slow response times in easier SNRs and speed response times in harder SNRs (see the General Discussion section for an explanation for why seeing the talker may slow response times in easy listening conditions).

Finally, I subsetted the data to include only the two SNRs tested in Experiment 1 (12 and -12 SNR) to enable more direct comparison with the finding from Experiment 1 that seeing the talker reduces dual-task costs in difficult listening conditions. These exploratory

<sup>8</sup> These effects also emerged when noise was coded numerically rather than categorically. See the Supplemental Materials.

<sup>9</sup> The interaction was also significant when SNR was coded numerically rather than categorically, as was the interaction between noise and modality on intelligibility. See the Supplemental Materials.

analyses followed the conventions outlined in Experiment 1. Neither the interaction ( $\chi^2_1 = 1.97, p = .16$ ) nor the effect of modality ( $\chi^2_1 = 0.03, p = .86$ ) was significant, but the effect of noise level was ( $\chi^2_1 = 17.16, p < .001$ ). Although nonsignificant in this second exploratory analysis, the significant interaction in the previous exploratory analysis and the mean response times in the four conditions of interest (i.e., audio-only and audiovisual response times at the easiest and hardest SNRs) were consistent with both our previous work assessing dual-task costs for audio-only and audiovisual speech (Brown & Strand, 2019a) and the results of Experiment 1. Indeed, in the easiest noise level, response times were 38 ms *slower* in the audiovisual (1,147 ms) than the audio-only (1,109 ms) condition, but in the hardest noise level, response times were 31 ms *faster* in the audiovisual (1,236 ms) than the audio-only (1,267 ms) condition.<sup>10</sup>

## General Discussion

This study aimed to evaluate the extent to which audiovisual relative to audio-only speech affects dual-task costs using isolated words, as well as measure the sensitivity, convergent validity, and psychometric properties of the novel dual-task paradigm across a range of background noise levels. In the following sections, I synthesize the results of the four experiments as they relate to these objectives and highlight directions for future research.

## Dual-Task Costs and Audiovisual Speech

Taken together, the results of the present study provide converging evidence that seeing the talker in addition to hearing their voice reduces the dual-task costs associated with processing speech in noise, but only when the listening conditions are sufficiently difficult that the visual signal substantially improves speech identification accuracy. In easy listening conditions, seeing the talker either had no effect on dual-task costs or actually *increased* dual-task costs. This trend emerged when study was conducted within subjects as well as between subjects and emerged when data were collected from undergraduates at Washington University in St. Louis as well as a more diverse sample of participants recruited online via Prolific. Crucially, these results are not attributable to inaudibility of the tones in louder background noise or to noise-induced cognitive interference that is unrelated to speech (see, e.g., Colle & Welsh, 1976); indeed, in two experiments using exactly the same noise and tone files, differences in response times to the tone task between the easy and hard noise levels robustly emerged when speech was present (Experiment 1) but not when it was absent (Experiment 2). Together, these results suggest that the slower response times in harder levels of background noise observed in Experiment 1, as well as the interaction between noise and sensory modality, were driven by increased cognitive demand associated with identifying the speech in difficult listening conditions.

These findings conceptually replicate previous work assessing dual-task costs for audio-only and audiovisual speech using a tactile dual-task paradigm (Brown & Strand, 2019a). That study used different speech stimuli, participant pools, and secondary tasks, and was conducted in-lab rather than remotely. Yet both studies found that any additional cognitive costs associated with processing audiovisual relative to audio-only speech in easy

conditions are greatly attenuated when the background noise is sufficiently loud (i.e., when accurate speech identification relies on both modalities). In this case, audiovisual speech *reduced* dual-task costs when the SNR was difficult.

Why might seeing the talker increase dual-task costs in easy listening conditions but decrease them in difficult conditions? In easy listening conditions, seeing the talker provides little or no intelligibility benefit. In this sense, the visual signal is superfluous, but additional resources must still be recruited to monitor both channels, process the (unnecessary) input, and/or integrate the inputs from the two modalities. Thus, when speech can be identified accurately from the auditory signal alone, the talking face may only serve as a distraction that diverts resources away from processing the auditory input (see Mishra et al., 2013).<sup>11</sup> When the level of the background noise is increased, however, spectral overlap between the masker and the target speech renders portions of the speech inaudible, which—according to the Ease of Language Understanding model—generates mismatches between the incoming acoustic input and phonological and lexical representations in memory. Additional cognitive resources must be recruited to resolve these mismatches, thereby diverting resources away from other simultaneous tasks (Rönneberg et al., 2008, 2010). However, some of this cognitive burden may be alleviated by visual cues provided by the talking face.

Given this theoretical account, the findings reported here are consistent with those reported in previous work. However, in the previous study using a vibrotactile dual-task paradigm and in-lab participants (Brown & Strand, 2019a), the main effect of modality was significant such that response times to the secondary task were *slower* in the audiovisual condition (holding SNR constant). In this study, however, the main effect of modality was nonsignificant in both Experiments 1 and 4. Why might the visual signal have slowed response times overall in the previous study, but not in this one? This pattern of results can be neatly accounted for by multiple resource theory: Given that the secondary tone classification task in the present study was auditory, according to multiple resource theory, this task competes for the same pool of auditory resources as the primary speech task, leading to greater interference than the vibrotactile secondary task used in Brown and Strand (2019a). Thus, although audiovisual speech reduces dual-task costs in both cases by relieving resources from the auditory pool (i.e., via reduced lexical competition), this reduction in costs is less pronounced in the previous study than in the present study because the competition for resources was more diffuse across three (auditory, visual, tactile)

<sup>10</sup> Note that an alternative method of evaluating the effects of seeing the talker on dual-task costs would be to match identification accuracy across modalities and compare audio-only and audiovisual response times across (matched) performance levels rather than SNRs. In this case, the conclusions of this “matched performance” method would suggest that seeing the talker always incurs a processing cost. However, a limitation of this approach is that it conflates SNR and modality; that is, matching performance across modalities requires that audiovisual speech be presented at a harder SNR than audio-only speech, so slower response times to the secondary task in audiovisual conditions may be driven by the louder background noise rather than the audiovisual modality, per se.

<sup>11</sup> It may also be that monitoring the inputs from the two modalities and integrating them into a unified percept occurs automatically, in which case the visual signal would not incur additional processing costs in easy conditions. Even if this were true, the visual signal would still not reduce dual-task costs in easy conditions, consistent with the results of this study and previous work (Brown & Strand, 2019a).



rather than two (auditory, visual) pools. In other words, the same reduction in lexical competition has a more pronounced effect in the present study because there are fewer resource pools available to begin with. As a result, the cost associated with monitoring inputs from multiple modalities is outweighed by the benefit in this study but not the previous one. This highlights the fact that for dual-task studies, task “difficulty” must be defined in terms of both speech identification accuracy (which is affected by the background noise level, talker and stimulus characteristics, etc.) and resource competition across multiple pools.

### Sensitivity, Convergent Validity, and Psychometric Properties of the Novel Tone Task

In recent years, several articles have argued for greater attention to measurement issues in psychology research (Flake & Fried, 2020). One of the issues that has received the most attention is the fact that researchers often create measures “on the fly” without justifying why they did so and without providing any evidence regarding the validity of the novel measure. Thus, given the novelty of the task, an additional goal of this study was to address this measurement issue and provide data enabling other researchers to further evaluate the psychometric properties of the new task to ensure that they are comfortable with the measurement tool before using it in their own research. Specifically, I established positive (Experiment 1) and negative (Experiment 2) control, evaluated the sensitivity of the task to changes in SNR (Experiment 3), provided evidence regarding the task’s convergent validity relative to another task in the listening effort literature (Experiment 3), and generated performance curves across a range of noise levels in audio-only and audiovisual conditions (Experiment 4).

#### Sensitivity

Experiment 3 demonstrated that the novel tone classification task was more sensitive to changes in SNR than the number classification task that has been used in previous research (Brown & Strand, 2019b; Picou & Ricketts, 2014; Sarampalis et al., 2009). The greater relative sensitivity of the tone task is unsurprising given the predictions of multiple resource theory: The tone task competes for the same pool of auditory perceptual resources as the speech task (along with the general pool of resources), whereas the number task only competes for resources in the general pool. Thus, as the speech identification task becomes more difficult, more resources from the auditory pool are allocated toward that task, leaving fewer available to complete the tone classification task. In contrast, although increasing the difficulty of the speech task certainly impairs performance on the number task (because the two tasks compete for resources in the general pool), this task is less affected by the difficulty of the speech task because the number task additionally has access to the visual pool of resources, which the (auditory-only) speech task does not. The high relative sensitivity of the tone task makes it a promising measurement tool for future researchers interested in the dual-task costs of speech processing.

#### Convergent Validity

In addition to assessing the sensitivity of the novel tone classification task to changes in SNR, Experiment 3 evaluated its convergent

validity by correlating performance on the task with performance on the number task at three SNRs. Results revealed that although single-task (i.e., no speech) response times were uncorrelated across the two tasks, when speech was introduced, correlations between dual-task response times increased substantially as the level of the background noise increased. On its face, it may appear unexpected that single-task response times were uncorrelated across the two tasks; both tasks rely on general processing speed, so it may be expected that individuals who can respond quickly to one task can also respond quickly to the other task. However, despite both being response time tasks, completing the two tasks in isolation (i.e., without speech) likely requires distinct sets of perceptual and cognitive abilities. The number task may rely on mathematical proficiency, the ability to inhibit the automatic response of pressing the key corresponding to the direction of the arrow (rather than, for trials with odd numbers, pressing the key corresponding to the opposite direction of the arrow), and working memory capacity (i.e., remembering which direction corresponds to odd vs. even numbers while simultaneously responding to the probe). In contrast, the tone task likely relies on auditory perceptual discrimination abilities as well as the ability to inhibit distracting information (i.e., momentarily ignore the speech when the tone occurs).<sup>12</sup>

Thus, although the single-task correlations suggest that the tone and number tasks largely rely on distinct sets of cognitive abilities, the dual-task correlations suggest a high degree of overlap in the mechanisms underlying secondary task processing while simultaneously identifying speech in noise. Indeed, if both tasks tap into a construct related to “effortful listening,” it would follow that the magnitude of the correlation between response times on the two tasks would increase with the difficulty of the speech task. In both cases, as more resources are allocated to the speech task—regardless of what those particular resources are—this leaves fewer remaining to complete the secondary task. Given that the two secondary tasks were paired with the same speech identification task, changes in the level of the background noise should similarly affect resource allocation to the primary task (though not identically because the secondary tasks themselves differ in resource requirements). Thus, although the sensitivities of the two tasks differ as a result of differences in resource overlap with the speech task, both tasks rely on pools of resources that become depleted with increases in background noise level, and this depletion of resources is reflected in the magnitude of the dual-task correlations across SNRs.

### Issues With Difference Scores and Proportional Dual-Task Costs

The correlations discussed above were calculated using *raw* (i.e., un-normalized) dual-task response times across the two tasks at each SNR; all three correlations were strong and statistically significant, and the trends were consistent with the predictions of multiple resource theory. However, when correlations were instead calculated using *derived* scores (i.e., difference scores and proportional dual-task costs), these values were greatly attenuated and in most cases were nonsignificant. A review of studies using proportional dual-task costs within the listening effort literature reveals that

<sup>12</sup> Note that this form of inhibitory control has been argued to be a separate ability from the ability to inhibit prepotent responses (see, e.g., Rey-Mermet et al., 2018), which is more relevant for the number task.

results tend to be particularly inconsistent both within and across studies when this derived score is used as the outcome variable (Gosselin & Gagné, 2011a, 2011b). These inconsistencies are likely driven by the poor reliability of difference and ratio (i.e., proportion) scores relative to the component scores.

Issues regarding the reliability of difference and ratio scores have a rich history dating back to at least 1941 (Cronbach, 1941; see also Lord, 1958). Intuitively, the reason for the poor reliability of difference and ratio scores is that derived scores like these have more sources of error than raw scores; each component of the derived score is subject to error, and this error compounds when the component scores are combined. In the case of the present study, the difference score derived by subtracting each participant's mean single-task response time (or their response time in the easier adjacent noise level) from their dual-task response time includes two sources of error: error associated with single-task response times and error associated with dual-task response times. Further manipulating the difference score by dividing it by the single-task response time introduces additional error in the resulting proportional dual-task cost score, which leads to even poorer reliability for this measure. Indeed, the results of the correlational analyses in Experiment 3 are perfectly consistent with what would be expected from classical measurement theory: The correlations involving raw scores were large and significant at all three SNRs, those involving difference scores were attenuated and only two were significant, and all of those involving proportional dual-task costs were nonsignificant.

### ***Pros and Cons of the Tone and Number Tasks***

The results of Experiment 3 suggest that a clear benefit of the novel tone classification task is its greater relative sensitivity to changes in noise than the number task, and therefore its potential utility for detecting smaller effects than other dual-task paradigms might be able to. Another benefit is that this task can be used in audiovisual speech research (unlike the visual number task) and can be implemented online (unlike the vibrotactile task), therefore enabling larger and more diverse participant samples than are possible in most research conducted in-lab. However, a drawback of the tone task is that accuracy at classifying the lengths of the tones was far lower (~54% across dual-task conditions in Experiment 3) than accuracy at classifying the parity of the numbers (~83% across dual-task conditions in Experiment 3), in part because the tone task consists of three response options whereas the number task consists of only two, but also in part because the tone task itself was more challenging. Regardless of the reason, this means that the same number of primary task trials produces fewer usable observations for the tone task than the number task (because response time trials are only included in the analyses when they are classified correctly), which may reduce statistical power and/or require that researchers collect data from more participants than they would need to if they used the number task.<sup>13</sup> Another potential drawback of the tone task is that the difference in speech identification accuracy between the moderate and hard noise levels was much larger for the tone task than the number task, suggesting that the tone task may have interfered with the listeners' ability to complete the primary speech task. Future researchers might consider making the tone task easier by adjusting the volume of the tones or making the tone lengths more distinct to increase the number of usable observations and mitigate interference with the speech task (see the Directions for Future Research section).

### ***Performance Curves***

Experiment 4 generated performance curves for accuracy and response times across nine SNRs for audio-only and audiovisual speech. These plots and the accompanying data provide researchers with valuable information that may help them determine the level of background noise that is most appropriate for the purposes of their own work. For example, if a researcher plans to conduct a study using this task and will present speech at two SNRs, they could use the results of Experiment 4 to determine the two SNRs that are likely to produce the largest difference in mean response times while keeping speech identification accuracy above a particular level (say, 75%) in a particular modality. Researchers may also use these performance curves to determine the intelligibility levels at which the task is likely to be most sensitive to changes in SNR, as well as the location of the asymptote at which even large changes in SNR are unlikely to affect response times. Finally, although Experiment 4 was not designed for these purposes, performance curves like these can also provide insight into whether matching SNR or instead matching speech identification accuracy (e.g., across participant groups or modalities of presentation) affects study outcomes.

### ***Directions for Future Research***

By making the stimuli and code for all four experiments publicly available, I hope that other researchers interested in the dual-task costs associated with speech processing (or indeed other forms of sensory and cognitive processing) will use the tone classification task to answer research questions beyond those addressed in the present study. Future work may also refine the task by varying the length and level of the tones to increase the number of usable observations, thereby improving the utility of the task for detecting subtle effects. Making the task easier in this way would also improve the precision of the mean response time estimates and therefore enable researchers to generate more precise performance curves. However, decreasing the difficulty of the task may reduce its sensitivity to changes in SNR or other experimental manipulations, so future researchers could attempt to identify combinations of parameters (e.g., tone length and level, frequency of the tones, width of the band of missing frequencies in the notched noise, etc.) that optimally balance measurement sensitivity and within-participant sample size (i.e., the number of trials completed by each participant). Researchers may also consider including trials in which the tone length was classified incorrectly to see if the results are reliant upon the decision to exclude them; if the results are consistent regardless of whether incorrect trials are excluded, this would obviate the need for making the task easier to increase the number of usable observations.

Finally, it is important to emphasize that the theoretical framework underlying my discussion of these experiments relies on multiple resource theory, but other explanations and theoretical approaches are certainly possible. An issue with any resource-related account of dual-task interference is that the precise resources being consumed are often unspecified or unknown. When a researcher uses a recall paradigm to assess listening effort, it is expected that working memory is at least one of the resources that affects an individual's overall

<sup>13</sup> Note, however, that this potential decrease in power may be at least partially offset by the tone task's greater relative sensitivity, which produces larger effect sizes.

performance on the task as well as the extent to which their performance is affected by the manipulation of interest (e.g., changes in SNR, modality, semantic constraint, etc.). But with dual-task paradigms in which the secondary task is not recall-based, it is less clear which particular resources are being consumed as task difficulty increases. This lack of an independent measure of resources leads to the tautological argument that is central to resource theories: The reason performance on secondary tasks becomes poorer in more difficult listening conditions is because these conditions require additional resources, which we know because performance on the secondary task became worse. The tautological nature of resource theory has been criticized since at least the early 1980s (see, e.g., Navon, 1984; Navon & Miller, 1987), and it is therefore important that researchers implementing dual-task paradigms consider alternative explanations for their findings before assuming that they must be driven by competition for limited resources.

The most common alternative theoretical account of dual-task interference involves what is often referred to as “cross-talk” or “outcome conflict” (Navon, 1984; Navon & Miller, 1987). This theory posits that dual-task interference arises not because tasks compete for the same input processing mechanisms, but because the output of one task interferes with the processing of the other task. Cross-talk models can certainly account for some dual-task findings, particularly when the two tasks are highly related (like two verbal tasks in which lexical activation elicited by one task may contaminate the lexical processing required of the other task). However, in order for cross-talk explanations to be viable, task difficulty must be defined by the degree of interference between outputs and concurrent processing; that is, increasing the difficulty of one task must also increase the extent to which the output of that task interferes with the processing of the other task. This is clearly not the case in the present study because there is no reason to expect that the output of the speech task (i.e., a lexical and semantic representation) provides greater interference with the processing of the tone task as the level of the background noise increases. Thus, although competition for limited resources from specific pools is the most feasible explanation for the results of the present study, the point remains that resource-based accounts can be tautological, and alternative explanations should be considered. To address this issue, future researchers could work to develop an independent measure of a “resource,” perhaps by conducting large-scale correlational studies with the goal of identifying the particular perceptual (e.g., auditory temporal processing) and cognitive (e.g., working memory) abilities that predict the degree of detriment on secondary tasks as speech tasks become more difficult.

## Conclusion

This study evaluated the effects of audiovisual speech on dual-task costs, which are thought to reflect online changes in listening effort and therefore have in-the-moment behavioral consequences for the listener. Using isolated words, I showed that seeing the talker improves the listener’s ability to complete a challenging simultaneous task when the background noise is loud, but may actually have the opposite effect when the listening conditions are sufficiently easy that the speech can be accurately identified without the visual signal. Additionally, by collecting data from large samples of online participants, this study provides valuable insight into the

behavioral consequences of speech identification in more naturalistic listening conditions and for listeners beyond the samples typically included in experimental psychology research (Henrich et al., 2010).

Finally, this work introduced a novel dual-task paradigm that can be implemented online in audiovisual conditions (which is not possible using other paradigms in the literature), established positive and negative control for this novel measure, assessed the sensitivity of the task to changes in SNR, and provided convergent validity evidence as well as data regarding the task’s psychometric properties across a range of listening difficulties. The materials necessary to implement the task and code for analyzing the data are publicly available at <https://osf.io/pqj8h/>. Thus, in addition to making theoretical contributions that are relevant to several literatures in cognitive psychology, this work produced a valuable tool that will enable researchers to answer theoretical questions related to the cognitive mechanisms supporting speech processing—as well as other forms of cognitive and perceptual processing—beyond the specific issues addressed here and without being limited by the necessity to conduct research in person.

## Constraints on Generality Statement (Simons et al., 2017)

The present study assessed how audiovisual speech affects dual-task costs across a wide range of background noise levels using online samples of young adults recruited via Prolific. The sample was not restricted to “native” English speakers (see Strand et al., 2024), so these findings are likely to generalize to the broader population of young adults based in the United States who are fluent in English (though likely not to individuals who are not fluent in the target language, in this case English). However, the sample included only young adults with self-reported normal hearing and normal or corrected-to-normal vision, so these findings may not extend to older adults, individuals with hearing loss, or cochlear implant users (though this is a fruitful avenue for future research).

One of the key findings is that the effects of seeing the talker on dual-task costs differed depending on the level of the background noise—or, more broadly, the difficulty of the task (see the General Discussion section)—so these results may only extend to tasks/listening conditions that approximate the difficulty of the task used here (i.e., the level/type of background noise, the type of speech materials [i.e., isolated words], the modality in which the secondary task is presented, and the difficulty of the secondary task and speech materials). However, the General Discussion section includes a detailed description of the expected relationship between task difficulty and the effects of audiovisual speech on dual-task costs, so these findings should extend broadly to other tasks and stimuli, but the direction of the effect of modality in a particular experimental condition will depend on task difficulty.

Finally, as discussed above, there are many ways to measure “listening effort,” and these various measures tap into different aspects of the construct. The results described here refer specifically to dual-task costs (and more specifically to dual-task costs when the secondary task is independent of the primary speech task) and therefore are unlikely to extend to other measures of listening effort, including subjective self-reports of effort, pupillometry, and recall measures.



## References

- Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, 40(5), 1084–1097. <https://doi.org/10.1097/AUD.0000000000000697>
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, 15(9), 839–843. <https://doi.org/10.1016/j.cub.2005.03.046>
- Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, 53(4), 1407–1425. <https://doi.org/10.3758/s13428-020-01501-5>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Basu Mallick, D., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, 22(5), 1299–1307. <https://doi.org/10.3758/s13423-015-0817-4>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Bernstein, L. E., Auer, E. T., Jr., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44(1–4), 5–18. <https://doi.org/10.1016/j.specom.2004.10.011>
- Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work, Environment & Health*, 16(1), 55–58. <https://doi.org/10.5271/sjweh.1815>
- Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (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>
- Brown, V. A., & Strand, J. F. (2019a). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition*, 2(1), Article 44. <https://doi.org/10.5334/joc.89>
- Brown, V. A., & Strand, J. F. (2019b). 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>
- Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44(4), 991–997. <https://doi.org/10.3758/s13428-012-0190-4>
- Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, 48(4), 687–697. <https://doi.org/10.1016/j.neuron.2005.09.029>
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1001–1010. <https://doi.org/10.1098/rstb.2007.2155>
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 17–31. [https://doi.org/10.1016/S0022-5371\(76\)90003-7](https://doi.org/10.1016/S0022-5371(76)90003-7)
- Cronbach, L. J. (1941). The reliability of ratio scores. *Educational and Psychological Measurement*, 1(1), 269–277. <https://doi.org/10.1177/001316444100100121>
- Desjardins, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, 34(3), 261–272. <https://doi.org/10.1097/AUD.0b013e31826d0ba4>
- 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(6), 600–610. <https://doi.org/10.1097/AUD.0000000000000028>
- Erber, N. P. (1972). Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research*, 15(2), 413–422. <https://doi.org/10.1044/jshr.1502.413>
- Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, 40(4), 481–492. <https://doi.org/10.1044/jshd.4004.481>
- Flake, J. K., & Fried, E. I. (2020). *Measurement schmeasurement: Questionable measurement practices and how to avoid them*. <https://doi.org/10.31234/osf.io/hs7wm>
- Fleming, J. T., Njoroge, J. M., Noyce, A. L., Perrachione, T. K., & Shinn-Cunningham, B. G. (2024). Sensory modality and information domain contribute jointly to dual-task interference between working memory and perceptual processing. *Imaging Neuroscience*, 2, 1–22. [https://doi.org/10.1162/imag\\_a\\_00130](https://doi.org/10.1162/imag_a_00130)
- 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](https://doi.org/10.1044/1092-4388(2009)08-0140)
- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, 106(6), 3578–3588. <https://doi.org/10.1121/1.428211>
- Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, 21. <https://doi.org/10.1177/2331216516687287>
- Gosselin, P. A., & Gagné, J.-P. (2011a). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, 50(11), 786–792. <https://doi.org/10.3109/14992027.2011.599870>
- Gosselin, P. A., & Gagné, J.-P. (2011b). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, 54(3), 944–958. [https://doi.org/10.1044/1092-4388\(2010\)10-0069](https://doi.org/10.1044/1092-4388(2010)10-0069)
- Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108(3), 1197–1208. <https://doi.org/10.1121/1.1288668>
- Grant, K. W., & Walden, B. E. (1996). Evaluating the articulation index for auditory-visual consonant recognition. *The Journal of the Acoustical Society of America*, 100(4), 2415–2424. <https://doi.org/10.1121/1.417950>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- 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](https://doi.org/10.1044/1092-4388(2002)046)
- Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 17(3), 259–273. <https://doi.org/10.1111/j.1469-8986.1980.tb00146.x>
- 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(3), 419–431. [https://doi.org/10.1044/2015\\_AJA-14-0058](https://doi.org/10.1044/2015_AJA-14-0058)
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.
- Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 46(2), 390–404. [https://doi.org/10.1044/1092-4388\(2003\)032](https://doi.org/10.1044/1092-4388(2003)032)
- Keidser, G., Best, V., Freeston, K., & Boyce, A. (2015). Cognitive spare capacity: Evaluation data and its association with comprehension of



- dynamic conversations. *Frontiers in Psychology*, 6, Article 597. <https://doi.org/10.3389/fpsyg.2015.00597>
- 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 masker. *Ear and Hearing*, 33(2), 291–300. <https://doi.org/10.1097/AUD.0b013e3182310019>
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, 50(1), 23–34. <https://doi.org/10.1111/j.1469-8986.2012.01477.x>
- Lord, F. M. (1958). The utilization of unreliable difference scores. *ETS Research Bulletin Series*, 1958(1), 150–152. <https://doi.org/10.1002/j.2333-8504.1958.tb00077.x>
- 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(2), 113–122. <https://doi.org/10.3766/jaaa.22.2.6>
- 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(1), 22–33. <https://doi.org/10.1080/02724980443000151>
- McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, 54(2), 193–203. <https://doi.org/10.1111/psyp.12772>
- McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021). Older adults show a more sustained pattern of effortful listening than young adults. *Psychology and Aging*, 36(4), 504–519. <https://doi.org/10.1037/pag0000587>
- 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(7), 433–445. <https://doi.org/10.3109/14992027.2014.890296>
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *bioRxiv*. <https://doi.org/10.1101/2020.07.21.214395>
- Mishra, S., Lunner, T., Stenfelt, S., Rönnerberg, J., & Rudner, M. (2013). Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers in Systems Neuroscience*, 7, Article 96. <https://doi.org/10.3389/fnsys.2013.00096>
- Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, 91(2), 216–234. <https://doi.org/10.1037/0033-295X.91.2.216>
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214–255. <https://doi.org/10.1037/0033-295X.86.3.214>
- Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 435–448. <https://doi.org/10.1037/0096-1523.13.3.435>
- Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnerberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, 52(7), 433–441. <https://doi.org/10.3109/14992027.2013.776181>
- Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, 365, 90–99. <https://doi.org/10.1016/j.heares.2018.05.003>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Pals, C., Sarampalis, A., & Baskent, 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\(2012/12-0074\)](https://doi.org/10.1044/1092-4388(2012/12-0074))
- Pals, C., Sarampalis, A., van Rijn, H., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America*, 138(3), EL187–EL192. <https://doi.org/10.1121/1.4929614>
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1), 593–608. <https://doi.org/10.1121/1.412282>
- 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>
- Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, 20(3), 241–248. <https://doi.org/10.1080/14640746808400158>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. <https://cran.r-project.org/bin/windows/base/old/4.2.2/>
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>
- Rönnerberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, 47(Suppl. 2), S99–S105. <https://doi.org/10.1080/14992020802301167>
- Rönnerberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health*, 12(49), 263–269. <https://doi.org/10.4103/1463-1741.70505>
- Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnerberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, 23(8), 577–589. <https://doi.org/10.3766/jaaa.23.7.7>
- 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\(2009/08-0111\)](https://doi.org/10.1044/1092-4388(2009/08-0111))
- Seeman, S., & Sims, R. (2015). Comparison of psychophysiological and dual-task measures of listening effort. *Journal of Speech, Language, and Hearing Research*, 58(6), 1781–1792. [https://doi.org/10.1044/2015\\_JSLHR-H-14-0180](https://doi.org/10.1044/2015_JSLHR-H-14-0180)
- Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85(1), 140–156. <https://doi.org/10.1080/03637751.2017.1342043>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>
- Slote, J., & Strand, J. F. (2016). Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods*, 48(2), 553–566. <https://doi.org/10.3758/s13428-015-0599-7>
- Sommers, M. S., & Phelps, D. (2016). Listening effort in younger and older adults: A comparison of auditory-only and auditory-visual presentations. *Ear and Hearing*, 37(1), 62S–68S. <https://doi.org/10.1097/AUD.0000000000000322>
- Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, 26(3), 263–275. <https://doi.org/10.1097/00003446-200506000-00003>
- Strand, J. F., Brown, V. A., & Barbour, D. L. (2020). Talking points: A modulating circle increases listening effort without improving speech

- recognition in young adults. *Psychonomic Bulletin & Review*, 27(3), 536–543. <https://doi.org/10.3758/s13423-020-01713-y>
- Strand, J. F., 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(6), 1463–1486. [https://doi.org/10.1044/2018\\_JSLHR-H-17-0257](https://doi.org/10.1044/2018_JSLHR-H-17-0257)
- Strand, J. F., Brown, V. A., Sewell, K., Lin, Y., Lefkowitz, E., & Saksena, C. G. (2024). Assessing the effects of “native speaker” status on classic findings in speech research. *Journal of Experimental Psychology: General*, 153(12), 3027–3041. <https://doi.org/10.1037/xge0001640>
- 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(4), 169–188. <https://doi.org/10.1080/25742442.2021.1903293>
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, 12(6), 462–466. <https://doi.org/10.1111/1467-9280.00386>
- Sumby, W. H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215. <https://doi.org/10.1121/1.1907309>
- Tomar, S. (2006, June 2). Converting video formats with FFmpeg. *Linux Journal*. <https://www.semanticscholar.org/paper/a439b564dbd8fc8a96bdb4f6d39ada749220f7dd>
- Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and Hearing*, 28(5), 656–668. <https://doi.org/10.1097/AUD.0b013e31812f7185>
- Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, 31(4), 380–389. <https://doi.org/10.1037/pag0000094>
- Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M. S., & Hale, S. (2011). Cross-modal enhancement of speech detection in young and older adults: Does signal content matter? *Ear and Hearing*, 32(5), 650–655. <https://doi.org/10.1097/AUD.0b013e31821a4578>
- Wagner, A. E., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, 7, Article 398. <https://doi.org/10.3389/fpsyg.2016.00398>
- Walden, B. E., Prosek, R. A., & Worthington, D. W. (1974). Predicting audiovisual consonant recognition performance of hearing-impaired adults. *Journal of Speech and Hearing Research*, 17(2), 270–278. <https://doi.org/10.1044/jshr.1702.270>
- Weisz, N., & Schlittmeier, S. J. (2006). Detrimental effects of irrelevant speech on serial recall of visual items are reflected in reduced visual N1 and reduced theta activity. *Cerebral Cortex*, 16(8), 1097–1105. <https://doi.org/10.1093/cercor/bhj051>
- Wickens, C. D. (1981). *Processing resources in attention, dual task performance, and workload assessment*. <https://www.semanticscholar.org/paper/dc27a217e9b1c52d77bcb867bb698e5bf7defb3a>
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449–455. <https://doi.org/10.1518/001872008X288394>
- Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7), 2064–2072. <https://doi.org/10.3758/s13414-017-1361-2>
- 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. (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>

Received September 27, 2023

Revision received May 29, 2025

Accepted June 30, 2025 ■