Dual-Task Paradigm Measures of Listening Effort: To Include or Not to Include Secondary Task Responses with Incorrect Primary Task Responses

Haiping Huang, Au.D.,[1] Ilze Oosthuizen, Ph.D.,[2,3] and Erin M. Picou, Au.D., Ph.D.[1,4]

ABSTRACT

Response time–based dual-task paradigms are commonly adopted to measure behavioral listening effort. Most extant studies used an all-response approach that included secondary task responses under both correct and incorrect primary task responses during analysis. However, evidence supporting this strategy is limited. Therefore, the current study investigated the potential differences between including all responses versus only including correct responses. Data from two previous studies were reanalyzed. Experiment 1 included 16 listeners and used a dual-task paradigm to examine the effect of introducing background noise on listening effort. Experiment 2 included 19 participants and used a different dual-task paradigm to examine the effect of reverberation and loudspeaker-to-listener distance on listening effort. ANOVA results obtained using both analysis approaches were compared. The all-response and correct-only approaches revealed similar results. However, larger effect sizes and an additional main effect were found with the all-response approach. The current study supports the use of an all-response approach due to its greater sensitivity to changes in behavioral listening effort. However, a correct-only approach could be utilized to suit specific study purposes.

KEYWORDS: listening effort, dual-task paradigm

1Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, Tennessee; 2Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa; 3Virtual Hearing Lab, University of Colorado School of Medicine, Aurora, Colorado; 4Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, Tennessee.

Address for correspondence: Haiping Huang, Au.D., Department of Hearing and Speech Sciences, Vanderbilt University, 1215 21st Avenue S, Room 8310, Nashville,TN 37232 (e-mail: haiping.huang@vanderbilt.edu).

Listening is not always easy. Effortful listening takes place whenever the person needs to deliberately allocate cognitive resources to a task of listening,[1] such as when listening in background noise,[2–4] in reverberation,[5–8] and to accented speech.[9] The understanding of listening effort is important because sustained increases in effort can result in mental fatigue[10,11] and may even lead to stress-related sick leave.[12] There are a variety of approaches to measure listening effort, including subjective reports, physiologic measures, memory paradigms, and response time paradigms.[13–15] One common approach reported in the literature is a dual-task paradigm, where participants perform a primary task (often speech recognition) and a simultaneous secondary task (for a review, see Gagne et al[16]).

By assuming humans have finite cognitive capacity,[17] the dual-task paradigm involves execution of two tasks simultaneously. While speech recognition mostly comprises the primary task, a great variety of activities have been utilized in the secondary task, including memory recall, response to probe lights, word categorization, and many others.[13,14,16] The change in performance (accuracy or speed) during the secondary task is considered an indication of reallocated cognitive resources. For example, in a word categorization response time paradigm, Picou and Ricketts[18] asked listeners to repeat the word they heard and press a button if the word can be a noun. Response time was the amount of time passed by between the start of the word presentation and button press in the secondary task. By assuming a limited pool of cognitive resources, longer response time/slower processing speed indicated more effort was allocated to the primary task.

It is noteworthy that the authors included all response time data in their analysis, no matter if the participant correctly or incorrectly repeated back the speech during primary task. We will refer to this approach as "all-response" approach for the rest of this article. This strategy of including all secondary task responses, independent of the primary task performance, is contrary to a "correct-only" approach that some investigators have adopted, where response times are excluded when the primary task responses were incorrect.[19,20] The argument supporting a correct-only approach is based on the idea that, unless a participant correctly repeats the speech in the primary task, they are not actively participating in the primary task and thus their response times might not appropriately reflect their effort.[21–23] Consistent with this concern, Wu et al[24] demonstrated that response times decreased (i.e., indicating less effort) when signal-to-noise ratio (SNR) became too challenging. Thus, response times that are very short might confound an estimation of effort during the speech recognition task. In other words, excessive guessing or giving up during trials when speech recognition becomes extremely difficult and frustrating could contaminate the dataset.

Yet, most studies in the extant literature have adopted an all-response approach.[2,3,6,13,24–26] See Table 1 for a brief review. This approach is supported by the current frameworks and models on listening effort. The Framework for Understanding Effortful Listening (FUEL) described that, whenever the incoming signal is distorted or degraded, listeners are more aroused and therefore the policy of mental resource allocation is changed so that more effort can be exerted to recognize the signal of interest.[1] Similarly, the Ease of Language Understanding (ELU) model argues that a mismatch between speech input and long-term memory storage will tax the working memory capacity.[27] Therefore, both FUEL and ELU imply that effort is most prominent when listeners have trouble identifying the auditory input. It is reasonable to deduce that perhaps the response times that correspond to incorrect speech recognition are of great importance and thus should not be excluded.

In support of this argument, Hsu et al[28] argued in favor of an all-response approach, highlighting that instances of incorrect speech recognition will systematically increase with increased background noise levels. Mishearing the speech in the primary task will be a natural consequence of listening in a challenging environment. Excluding response times when the primary task was incorrect might underrepresent the amount of effort a participant is allocating to the speech recognition task. For example, if an experimenter is targeting a 70% speech recognition performance level for the primary task, but the only response times

Table 1 Summary table of studies using different approaches

Authors and effects researched	Primary (P) and secondary (S) tasks	Approach
Baer et al[19]: Effects of signal processing, SNR	P: Four-word sentence closed set recognition S: RT for judgment of P task sentence as sensible or silly	Correct-only
Picou et al[2]: Effects of modality and amplification	P: Word recognition (monosyllabic) S: RT to visual probe	All-response
Picou and Ricketts[18]: Effects of secondary task on dual-task paradigm sensitivity to listening effort; also modality and hearing status	P: Word recognition S1: Simple—RT to visual probe S2: Complex—RT to visual stimulus S3: Semantic—RT on semantic category judgment	All-response
Picou et al[6]: Effects of noise and reverberation	P: Monosyllabic word recognition S: RT on word-class judgment	All-response
Picou et al[26]: Effects of modifying secondary task in dual-task paradigm; child–adult differences	P: Word recognition S1: Simple—RT to visual probe S2: Complex—RT to complex visual probe S3: Deep—RT to word categorization	All-response
Sarampalis et al[3]: Effects of signal processing, low or high context, SNR	Exp 2 P: Sentence recognition S: RT to complex visual RT task	All-response
Seeman and Sims[38]: Comparison of psychophysiological and dual-task on listening effort	Exp 2 P: Sentence recognition S: RT to complex visual RT task	All-response
Strand et al[13]: Evaluate the convergent validity and sensitivity of commonly used measures of listening effort and assess how scores on those tasks relate to cognitive and personality variables	Dual-task measure: P: Monosyllabic word recognition S1: Complex—RT to visual stimulus S3: Semantic—RT on semantic category judgment	All-response
Wu et al[25]: Effects of amplification by signal processing	Exp 2 P: Speech recognition S: RTs on complex visual RT task	All-response
Wu et al[24]: Characterized the psychometric functions that describe task performance in dual-task listening effort measures as a function of SNR	P: Speech recognition (HINT) S1: Simple visual-reaction time task S2: Hard—Stroop test paradigm	All-response

Abbreviations: HINT, hearing-in-noise test; P, primary task; RT, response time; S, secondary task; SNR, signal-to-noise ratio.

analyzed are for correct primary task trials, the effective performance level for the effort task would be 100%. Consequently, including all response times will give a more realistic indication of effort in the given listening situation, because it will reflect the targeted, rather than perfect, performance.

Few studies in the field have directly addressed the comparison of the all-response and correct-only approaches. As an aside, Hicks and Tharpe[20] stated in their data analysis that the numerical difference between correct-only and all-response times were only less than 14 ms, which was considered negligible. However, the authors did not do a formal comparison or present both sets of data. McGarrigle et al[29] took a similar approach, reporting only the all-response data set, but describing anecdotally that analyses with the correct-only data set revealed an identical pattern of results as an all-response approach. Therefore, the issue of which approach to take, correct-only or all-response, is yet unresolved. The approach that is most closely linked to the conceptual frameworks and models of listening effort (i.e., FUEL and ELU) and is the most sensitive to factors expected to affect effort would be the most desirable one (for similar arguments see Fisk et al[21]).

Purpose

The current study aimed to evaluate a methodological difference in the dual-task paradigm. Specifically, our goal was to determine whether we should treat response times collected under incorrect word recognition as natural listening consequences (i.e., all-response approach) or as data contaminants (i.e., correct-only approach). To achieve this, we reanalyzed two datasets from previous experiments in our laboratory. In Experiment 1, we reevaluated secondary task data from a study examining the effect of noise on listening effort, focusing on one paradigm and adults with normal hearing.[26] In Experiment 2, we reevaluated secondary task data from a study examining the effects of loudspeaker–listener distance and reverberation on behavioral listening effort.[30] Based on the nonsignificant difference in the results of different response approaches reported by others,[20,29] we hypothesized that using the all-response approach versus the correct-only approach in a dual-task paradigm would not alter the conclusions of these studies. The results of this study will provide empirical support that the exclusion criteria for secondary task response times will not affect study conclusions and would provide support for future studies to use either approach.

EXPERIMENT 1: METHODS

Detailed methodological descriptions are reported elsewhere[26] and are only summarized here. In brief, the goal of this experiment was to examine the effects of introducing background noise and of increasing secondary task complexity and depth of processing in a dual-task paradigm for listening effort testing in adults and children. For the current study, to facilitate comparison with Experiment 2, only the data from adults and the secondary task requiring deep processing were reanalyzed and discussed, specifically those completed in quiet and in background noise.

Participants

The project included 16 adults ($M = 25.4$ years, standard deviation [SD] = 3.3, range = 22–32 years). Participants had bilateral normal hearing sensitivity (<25 dB HL at audiometric octaves) indicated by hearing screening. They all reported normal middle ear function, and no otologic, cognitive, or neurogenic disorders. Testing was completed with the approval of Vanderbilt University's Institutional Review Board. All participants were compensated for their time.

Dual-Task Paradigms

The primary task involved word recognition of a female talker speaking monosyllable words (concrete nouns) and presented at a level of 65 dBA. The speech material consisted of eight lists of 25 words each; development of the words has been detailed previously.[2,18,31] Each word is approximately 1,700 ms long. Background noise, when present, was a four-talker babble, consisting of four female talkers reading passages from the Connected Speech Test (CST).[32,33] Each talker's voice was presented from one of four noise loudspeakers. Overall background noise level was 69 dBA, resulting in a −4 dB SNR. The participants were tested in quiet and in noise, with four-word lists in each in condition.

The secondary task was a categorical size judgment of the words presented. Listeners were asked to judge whether the word they heard was generally bigger in size than a basketball (e.g., pond, man). If so, they should touch a computer screen placed in front of them. This physical response was timed. There were nonprobe and probe trials in the secondary task. During nonprobe trials, words presented represented objects that are not bigger than a basketball (e.g., mouse, bean) and participants were asked to perform only the primary task of word recognition. During probe trials, words presented represented objects that can be bigger than a basketball and participants were asked to perform both tasks. A white square (6.5 × 6.5 cm) was present during both probe and non-probe trials and disappeared as soon as a participant touched the screen. The square appeared 500 ms after word onset. Response time indexed behavioral listening effort and was the time elapsed between the start of the word presentation and the participant's physical response.

Test Environment

Testing took place in a sound-attenuating test booth (4 m × 4.3 m × 2.7 m). The speech stimuli were routed to a loudspeaker (Tannoy System 600, Coatbridge, Scotland) located 1 m in front of the listener. The computer screen (Dell S2240T, Round Rock, TX) was placed on a table below the loudspeaker and served as the response system for recording secondary task responses. The four-talker babble, when present, was played through four loudspeakers (Definitive BP-2X, Definitive Technologies, Owings Mills, MD) located at 45, 135, 225, and 315 degrees, 1 m from the listener.

Data Analysis

For the purpose of the current study, primary task performance was summarized descriptively. Prior to statistical testing, response times less than 100 ms or longer than 3 SD from the mean were excluded. Note that this technique will be able to exclude most premature responses, but will not rule out all guessing attempts. Means were obtained from four trials for each condition. Then, response times (in ms) were analyzed using repeated-measures analysis of variance (ANOVA) with one within-participant factor, noise (quiet and noise). Two separate ANOVAs were conducted, once with an all-response approach and once with a correct-only approach to evaluate if the different exclusion criteria would result in a different pattern of statistical results or different conclusions. In both cases, Shapiro–Wilk tests of normality revealed the studentized residuals of response time data were normally distributed ($p > 0.05$). There were no outliers, as assessed by studentized residuals greater than ± 3 SD. Effect size was measured using partial eta squared. All analyses were completed with SPSS (v.26 IBM Corporation).

EXPERIMENT 1: RESULTS

Mean word recognition performance for the primary task in each condition is displayed in Table 2. Using both all-response and correct-only approaches, repeated measures ANOVAs were conducted to determine whether the introduction of background noise increased participants' listening effort, as indicated by prolonged response times. When correct-only responses were included, background noise resulted in a statistically significant increase in response times, F (1,15) = 36.51, $p < 0.001$, partial $\eta^2 = 0.700$. Mean response times are shown in Table 3. It is clear that response times were faster when listening in quiet compared to listening in noise, reflecting less listening effort in quiet.

When all responses were included, background noise also affected response times, F (1,15) = 72.359, $p < 0.001$, partial $\eta^2 = 0.828$. Mean response times are shown in Table 3. Response times were faster when listening in quiet compared to listening in noise, indicating reduced listening effort in quiet. These findings indicate both approaches presented similar outcomes; however, the effect size observed using the all-response approach was larger.

Table 2 Word recognition performance for all conditions in both experiments

	Experiment 1		Experiment 2			
	Quiet	Noise	Moderate reverberation		High reverberation	
			Inside CD	Outside CD	Inside CD	Outside CD
WR	100%	52%	87%	73%	73%	48%

Abbreviations: CD, critical distance; WR, word recognition.
Note: Word recognition for both experiments.

Table 3 Mean response times (ms) in Experiment 1

Response times (ms)	Quiet	Noise	Condition difference (ms)
Correct-only	1,353.9	1,557.6	203.7 (95% CI, 131.8–275.5, SD = 135)
All-response	1,353.9	1,595	241.1 (95% CI, 180.7–301.5, SD = 114)
Approach difference	0 (SD = 0.3)	37.4 (SD = 62.8)	

Abbreviations: CI, confidence interval; SD, standard deviation.
Note: CIs are indicated where significance testing was conducted, and significant result was obtained.

EXPERIMENT 2: METHODS

Experiment 2 was a part of a publication in progress.[30] The purpose of the experiment was to evaluate the effects of listener-to-loudspeaker distance and reverberation on behavioral listening effort for adults with normal hearing. Only a subset of the conditions is included in the current study.

Participants

Participants included 19 adults ($M = 31$ years, SD = 5.5, range = 19–40 years) with normal hearing sensitivity in both ears (i.e., hearing thresholds better than 25 dB HL at audiometric octaves from 250 to 8,000 Hz), as measured with standard audiometry. Participants denied history of chronic middle ear disease or neurologic disorder indicated by self-report. The Bamford–Kowal–Bench speech-in-noise test (BKB-SIN)[34] was administered bilaterally, through insert earphones at 70 dB HL to assess listener's speech recognition in noise abilities. The average BKB-SIN, SNR-50 score for listeners was 0.5 dB (range: −1 to 2 dB, SD = 0.96). Testing was completed with the approval of Vanderbilt University Medical Center's Institutional Review Board. All participants were compensated for their time.

Dual-Task Paradigms

Similar to Experiment 1, the primary task was monosyllable word recognition using the same female talker and a word categorization secondary task. Unique to Experiment 2, there were 8, approximately equally intelligible, 60-word lists; words were not only concrete nouns but could also be verbs or adjectives. Participants were tested with two lists of words in each test condition. The secondary task was to judge whether the word was a noun (rather than bigger than a basketball in Experiment 1). The presentation level of the speech was 65 dBA. The same background noise was a four-talker babble used in Experiment 1. The development for the speech and noise material was reported in previous literature.[18,31] All testing was completed in the presence of background noise with an individualized SNR. For each participant, the test SNR was obtained by subtracting 3 dB from participant's BKB-SIN SNR-50 scores. This SNR was kept constant across conditions within a participant's test battery. For example, if the participant's BKB-SIN SNR-50 score was 1 dB, we used a −2 dB SNR for all of his testing. The goal was to target approximately 84% speech recognition performance in the easiest condition (inside critical distance, moderate reverberation) for all participants.

Test Environment

A reverberant room (5.5 m × 6.5 m × 2.25 m) was used for the experiment. Test conditions varied by reverberation time (i.e., moderate and high) and listener-to-loudspeaker distances (i.e., inside and outside critical distance). Frequency-specific reverberation times (RT_{30}) are shown in Table 4. To achieve the two reverberation times, we adjusted the reverberant characteristic of the room using acoustic blankets (Sound Stopper 4 × 8) and floor carpeting. We removed all acoustic blankets in the room during high reverberant ($RT_{30} = 1{,}223$ ms) conditions, while leaving eight acoustic blankets on the walls and the ceiling during moderate reverberant ($RT_{30} = 469$ ms) conditions. For the listener-to-loudspeaker distances, we used the functional formula from Peutz[35] and calculated the critical distance for the two

Table 4 Reverberation time (RT_{30}, s) as a function of frequency (Hz)

Frequency (Hz)	630	1,250	2,500	5,000	10,000	Mean
Moderate (RT_{30}, s)	0.42	0.46	0.48	0.51	0.41	0.47
High (RT_{30}, s)	1.55	1.47	1.01	0.86	0.51	1.22

reverberant conditions based on the room size. The two loudspeaker-to-listener distances we designated were 4 m (outside critical distance) and 1.25 m (inside critical distance) across both reverberation conditions. Based on the parameters described earlier, four test conditions were included in the experiment, namely: (1) moderate reverberation/inside critical distance, (2) moderate reverberation/outside critical distance, (3) high reverberation/inside critical distance, and (4) high reverberation/outside critical distance.

During listening effort testing, speech stimuli were presented from a loudspeaker (Tannoy 600A) positioned directly in front of the listener. This speech loudspeaker, along with four noise loudspeakers (Tannoy 600), had fixed positions in all conditions while the listener's location was changed. The participant was positioned at 1.25 or 4 m from the speech loudspeaker, for the inside and outside critical distance conditions, respectively, in both reverberation conditions. The four noise loudspeakers were placed at four corners of the room, used to present the four-talker babble. When the listener was inside critical distance from the speech loudspeaker, the four noise loudspeakers were at a distance of 2.5, 2.5, 3.8, and 3.8 m from the listener. When the listener was outside critical distance from the speech loudspeaker, the four noise loudspeakers were at a distance of 4.6, 2.1, 2.1, and 4.6 m from the listener.

Data Analysis

For the current study, primary task performance was summarized descriptively. As in Experiment 1, prior to statistical analysis, response times that were less than 100 ms or longer than 3 SD from the mean were excluded. Means were obtained from two repetitions in each condition. Response times (ms) were analyzed with repeated-measures ANOVA, once using an all-response approach and once using a correct-only approach. Each ANOVA included two within-participant factors, reverberation (moderate and high) and listener-to-talker distance (inside and outside critical distance). Shapiro–Wilk test of normality revealed the studentized residuals of response time data were not normally distributed when all-response approach was used, but were normally distributed when correct-only approach was used. All-response data were transformed with an inverse function (transformed_rt = 1/original_rt), and residuals were then distributed within normal limits. There were no outliers, as assessed by studentized residuals greater than ± 3 SD. Effect size was measured using partial eta squared. All analyses were completed with SPSS (v.26 IBM Corporation).

EXPERIMENT 2: RESULTS

Mean word recognition performance for the primary task in each condition is displayed in Table 2. Using a correct-only approach, analysis revealed significant main effects for listener-to-loudspeaker distance, F (1, 18) = 6.631, $p = 0.019$, partial $\eta^2 = 0.269$. Mean response times are shown in Table 5. Response times were faster when listening inside the critical distance of the room compared to outside the critical distance. There was no significant main effect of reverberation, F (1, 18) = 4.005, $p = 0.061$, partial $\eta^2 = 0.182$, or significant interaction between reverberation and listener-to-loudspeaker distance, F (1, 18) = 0.063, $p = 0.448$, partial $\eta^2 = 0.032$.

Using an all-response approach, analysis revealed significant main effects for reverberation, F (1, 18) = 9.459, $p = 0.007$, partial $\eta^2 =$ 0.34. Main effects were also significant for listener-to-loudspeaker distance, F (1, 18) = 18.978, $p < 0.001$, partial $\eta^2 = 0.51$. Mean response times are shown in Table 5. Response times were faster in the moderate reverberation condition compared to the high reverberation

Table 5 Mean response times (ms) in Experiment 2

	Moderate reverberation		High reverberation	
Response times (ms)	Inside CD	Outside CD	Inside CD	Outside CD
Correct-only	2,218.5	2,274.9	2,262.2	2,385.9
All-response	2,252.4	2,334	2,308.7	2,514.1
Approach difference	33.9 (SD = 50.3)	59.1 (SD = 83.2)	46.5 (SD = 77)	128.2 (SD =151)
Condition difference	Moderate vs. high reverberation		Inside vs. outside CD	
Correct-only	77 (95% CI, −3.9 to 158.6)		90 (95% CI, 16.6–163.6)	
All-response	2.4×10^{-5} (95% CI, 7.5×10^{-6} to 4×10^{-5}). Pretransformed: 117		2.5×10^{-5} (95% CI, 1.3×10^{-5} to 3.7×10^{-5}). Pretransformed: 145	

Abbreviations: CD, critical distance; CI, confidence interval; SD, standard deviation.
Notes: CIs are indicated where significance testing was conducted. Pre–data transformation mean values are provided for condition differences under all-response approach, for ease of interpretation of readers.

condition and when listening inside the critical distance of the room compared to outside the critical distance. There was no significant interaction between reverberation and listener-to-loudspeaker distance, $F(1, 18) = 1.77$, $p = 0.2$, partial $\eta^2 = 0.09$. Collectively, this set of findings indicates that the two approaches yielded different outcomes. In contrast with the findings from Experiment 1, we observed more significant main effects in addition to larger effect sizes with Experiment 2, indicating an all-response approach may be a more sensitive analytic approach using this dual-task paradigm.

GENERAL DISCUSSION

The purpose of this study was to determine whether we should treat response times collected under incorrect word recognition as natural listening consequences (i.e., all-response approach) or as data contaminants (i.e., correct-only approach). The approach that is the most sensitive to factors expected to affect effort would be the most desirable one. Therefore, we reanalyzed two datasets from previous experiments in our laboratory reevaluating the effects of background noise (Experiment 1) and listener-to-loudspeaker distance and reverberation (Experiment 2). Although most studies in the extant literature have used an all-response approach to dual-task paradigms,[2,3,6,13,24–26] the rationale behind a correct-only approach is potentially twofold. First, when listening becomes very challenging, participants may experience cognitive overload, causing them to actively disengage from the task.[24] As a result, listeners may exert less effort under unfavorable conditions (i.e., low SNR, high reverberation, etc.). Therefore, response times collected using an all-response approach could potentially be much shorter than those using a correct-only approach, due to listener's excessive disengagement during the primary task and guessing during the secondary task. Second, some might argue that mental effort exerted during incorrect word recognition is negligible.

Our findings from both experiments, however, counter these two arguments. We saw that mean response times using the correct-only approach were in fact shorter than mean responses using an all-response approach (see Tables 3 and 5). Additionally, larger effect sizes (background noise in Experiment 1 and listener-to-loudspeaker distance in Experiment 2), as well as an additional significant main effect (reverberation time in Experiment 2), were observed when we used an all-response approach. This combination of findings suggests that instead of guessing confounding the analysis, the all-response approach was more sensitive to factors of interest and resulted in larger effect sizes than the correct-only approach.

There are a few reasons why the correct-only approach ended up yielding shorter responses times. First, it might be that our participants stayed engaged in the primary task because performance levels were around 50%

or above. According to the performance-intensity function provided by Wu et al,[24] disengagement and faster response times is more apparent around 40% performance on the primary task. Thus, the participants in our studies were less likely to disengage, because their performance was relatively good. Second, it could be that unsuccessful word recognitions contained more processing than successful ones, consistent with the ELU model.[27] The model states that an unsuccessful match of incoming speech signal and long-term memory will trigger loops of explicit processing that involves interference of the working memory system (e.g., attention shifting, inference making, semantic integration), resulting in a relatively slower process.[27] It could be that multiple trials of retrieval failure-cognitive compensation-retrieval reattempt occurs until either the retrieval attempt succeeds, or the listener gives up. If this argument holds, then excluding effort related to incorrect word recognition may be detrimental since we are ignoring a listener's hard work solely based on the outcome. In other words, processing and allocation of mental resource are present in any listening task,[36] even if a participant does not correctly repeat the primary task speech

Regarding all-response approach showing bigger effect size in both experiments and the additional main effect from reverberation in Experiment 2, this could be due to insufficient data points under correct-only approach. Findings from a previous study has demonstrated the importance of having enough observation counts to interpret differences in response times.[37] This is especially significant when testing was conducted under challenging SNRs where word recognition performance is compromised. For example, Experiment 1 yielded 52% mean word recognition when background noise was present. With correct-only approach, where word recognition performance for the response time trials is 100%, we collected 467 response time data points. In comparison, we obtained 670 data points with all-response approach. In Experiment 2, mean word recognition was 48% when listeners were tasked under high reverberation, outside critical distance. In this condition alone, response time counts dropped from 963 to 655 when correct-only approach was used.

An all-response approach during analysis of secondary task response times in a dual-task paradigm may have some theoretical advantages over correct-only approach. However, it is noteworthy that a correct-only approach could certainly be used to suit some study purposes. One potential use of this approach is to investigate listening effort under conditions of comparable word recognition performance levels. With a correct-only approach, word recognition performance remains at 100%; therefore ,if behavioral listening effort is at the same time significantly different across conditions, it is likely that word recognition performance was not driving behavioral listening effort. Similarly, correct-only approach will be useful when we are trying to gage whether perceived listening effort is better related to word recognition or behavioral effort. Since word recognition stays at 100% with a correct-only approach, if perceived effort stays relatively stable across different listening conditions, then it would be reasonable to conclude that perceived effort is driven by word recognition during listening.

CONCLUSION

The current study aimed to determine whether including response time data associated with both correct and incorrect word recognition would jeopardize the interpretation of dataset. We reanalyzed data from two previous experiments. The results demonstrated that an all-response approach yielded larger effect sizes in both experiments and an additional significant main effect in Experiment 2. Therefore, the current study supports the use of an all-response approach due to its greater sensitivity to behavioral listening effort. However, we also recognize that "correct-only approach" can be utilized to suit specific study purposes.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). Ear Hear 2016;37(Suppl 1):5S–27S

2. Picou EM, Ricketts TA, Hornsby BWY. How hearing aids, background noise, and visual cues influence objective listening effort. Ear Hear 2013; 34(05):e52–e64
3. Sarampalis A, Kalluri S, Edwards B, Hafter E. Objective measures of listening effort: effects of background noise and noise reduction. J Speech Lang Hear Res 2009;52(05):1230–1240
4. Ohlenforst B, Wendt D, Kramer SE, Naylor G, Zekveld AA, Lunner T. Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. Hear Res 2018;365:90–99
5. Prodi N, Visentin C. Impact of background noise fluctuation and reverberation on response time in a speech reception task. J Speech Lang Hear Res 2019;62(11):4179–4195
6. Picou EM, Gordon J, Ricketts TA. The effects of noise and reverberation on listening effort in adults with normal hearing. Ear Hear 2016;37(01):1–13
7. Peng ZE, Wang LM. Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. J Speech Lang Hear Res 2019;62(04): 1068–1081
8. Sato H, Morimoto M, Wada M. Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons. J Acoust Soc Am 2012;131(06):4596–4605
9. Van Engen KJ, Peelle JE. Listening effort and accented speech. Front Hum Neurosci 2014;8:577
10. Hornsby BWY. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. Ear Hear 2013;34(05):523–534
11. Wang Y, Naylor G, Kramer SE, et al. Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. Ear Hear 2018;39(03): 573–582
12. Kramer SE, Kapteyn TS, Houtgast T. Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. Int J Audiol 2006;45(09):503–512
13. Strand JF, Brown VA, Merchant MB, Brown HE, Smith J. Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. J Speech Lang Hear Res 2018;61(06):1463–1486
14. McGarrigle R, Munro KJ, Dawes P, et al. Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. Int J Audiol 2014;53(07):433–440
15. Alhanbali S, Dawes P, Millman RE, Munro KJ. Measures of listening effort are multidimensional. Ear Hear 2019;40(05):1084–1097
16. Gagné J-P, Besser J, Lemke U. Behavioral assessment of listening effort using a dual-task paradigm: a review. Trends Hear 2017;21:2331216516687287
17. Kahneman D. Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall1973
18. Picou EM, Ricketts TA. The effect of changing the secondary task in dual-task paradigms for measuring listening effort. Ear Hear 2014;35(06): 611–622
19. Baer T, Moore BC, Gatehouse S. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times. J Rehabil Res Dev 1993;30(01):49–72
20. Hick CB, Tharpe AM. Listening effort and fatigue in school-age children with and without hearing loss. J Speech Lang Hear Res 2002;45(03):573–584
21. Fisk AD, Derrick WL, Schneider W. A methodological assessment and evaluation of dual-task paradigms. Curr Psychol Res Rev 1986;5(04): 315–327
22. Watter S, Logan GD. Parallel response selection in dual-task situations. Percept Psychophys 2006;68 (02):254–277
23. Yamaguchi M, Logan GD, Bissett PG. Stopping while going! Response inhibition does not suffer dual-task interference. J Exp Psychol Hum Percept Perform 2012;38(01):123–134
24. Wu Y-H, Stangl E, Zhang X, Perkins J, Eilers E. Psychometric functions of dual-task paradigms for measuring listening effort. Ear Hear 2016;37(06): 660–670
25. Wu Y-H, Aksan N, Rizzo M, Stangl E, Zhang X, Bentler R. Measuring listening effort: driving simulator versus simple dual-task paradigm. Ear Hear 2014;35(06):623–632
26. Picou EM, Charles LM, Ricketts TA. Child–adult differences in using dual-task paradigms to measure listening effort. Am J Audiol 2017;26(02):143–154
27. Rönnberg J, Lunner T, Zekveld A, et al. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front Syst Neurosci 2013;7:31
28. Hsu BC-L, Vanpoucke F, van Wieringen A. Listening effort through depth of processing in school-age children. Ear Hear 2017;38(05): 568–576
29. McGarrigle R, Gustafson SJ, Hornsby BWY, Bess FH. Behavioral measures of listening effort in school-age children: Examining the effects of signal-to-noise ratio, hearing loss, and amplification. Ear Hear 2019;40(02):381–392
30. Huang H, Ricketts TA, Hornsby BWY, Picou EM. Effects of critical distance and reverberation

on listening effort in adults. J Speech Lang Hear Res 2022;65(12):4837–4851
31. Picou EM, Ricketts TA, Hornsby BWY. Visual cues and listening effort: individual variability. J Speech Lang Hear Res 2011;54(05):1416–1430
32. Cox RM, Alexander GC, Gilmore C. Development of the connected speech test (CST). Ear Hear 1987;8(5, Suppl):119S–126S
33. Cox RM, Alexander GC, Gilmore C, Pusakulich KM. Use of the connected speech test (CST) with hearing-impaired listeners. Ear Hear 1988;9(04): 198–207
34. Etymotic R. Bamford-Kowal-Bench Speech-in-Noise Test (Version 1.3) Manual, [Audio CD]. Elk Grove Village, IL: Etymotic Research; 2005
35. Peutz V. Articulation loss of consonants as a criterion for speech transmission in a room. J Audio Eng Soc 1971;19(11):915–919
36. Lemke U, Besser J. Cognitive load and listening effort: concepts and age-related considerations. Ear Hear 2016;37(Suppl 1):77S–84S
37. Brysbaert M, Stevens M. Power analysis and effect size in mixed effects models: a tutorial. J Cogn 2018;1(01):9
38. Seeman S, Sims R. Comparison of psychophysiological and dual-task measures of listening effort. J Speech Lang Hear Res 2015;58(06):1781–1792