

Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm: A Review

Trends in Hearing
2017, Vol. 21: 1–25
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516687287
journals.sagepub.com/home/tia



Jean-Pierre Gagné¹, Jana Besser², and Ulrike Lemke²

Abstract

Published investigations ($n = 29$) in which a dual-task experimental paradigm was employed to measure listening effort during speech understanding in younger and older adults were reviewed. A summary of the main findings reported in the articles is provided with respect to the participants' age-group and hearing status. Effects of different signal characteristics, such as the test modality, on dual-task outcomes are evaluated, and associations with cognitive abilities and self-report measures of listening effort are described. Then, several procedural issues associated with the use of dual-task experiment paradigms are discussed. Finally, some issues that warrant future research are addressed. The review revealed large variability in the dual-task experimental paradigms that have been used to measure the listening effort expended during speech understanding. The differences in experimental procedures used across studies make it difficult to draw firm conclusions concerning the optimal choice of dual-task paradigm or the sensitivity of specific paradigms to different types of experimental manipulations. In general, the analysis confirmed that dual-task paradigms have been used successfully to measure differences in effort under different experimental conditions, in both younger and older adults. Several research questions that warrant further investigation in order to better understand and characterize the intricacies of dual-task paradigms were identified.

Keywords

listening effort, dual-task paradigm, speech recognition, speech understanding, cognitive resources

Date received: 27 July 2016; revised: 6 December 2016; accepted: 8 December 2016

Introduction

In clinical settings, speech understanding is typically measured by calculating the proportion of keywords that can be identified correctly under a given listening condition (e.g., in quiet or in noise). One aspect of speech understanding that is undervalued is listening effort. Listening effort refers to 'the amount of processing resources (perceptual, attentional, cognitive, etc.) allocated to a specific auditory task, when the task demands are high (adverse listening conditions) and when the listener strives to reach a high-level of performance on the listening task. Under ideal listening conditions, listening to speech is relatively effortless (Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008). Processing speech may become more effortful when the quality of the signal is degraded (e.g., due to noise or if the listener has hearing loss), when the language structure used is complex, or when the content of the message is less familiar.

There is no direct relationship between performance level in terms of proportion (or percent) of correct responses obtained on a listening task and listening effort. A person may obtain the same proportion of

correct responses on two different tasks but report that performing one task was substantially more effortful than the other one. For example, a listener may be able to fully understand a spoken message in a challenging background noise but the amount of listening effort required to process the message in this situation may be considerably greater than when the same signal is processed in a quiet background setting. Likewise, persons with hearing loss often report that in some environments (especially noisy ones) listening requires substantially more

¹École d'orthophonie et d'audiologie, Chaire de recherche Caroline Durand en audition et vieillissement, Centre de recherche de, L'Institut universitaire de gériatrie de Montréal, Université de Montréal, Québec, Canada

²Phonak's research program in Cognitive & Ecological Audiology, Department of Science & Technology, Phonak AG, Science and Technology – CEA (Q480), Stäfa, Switzerland

Corresponding author:

Jean-Pierre Gagné, École d'orthophonie et d'audiologie, Chaire de recherche Caroline Durand en audition et vieillissement, Centre de recherche de, L'Institut universitaire de gériatrie de Montréal, Université de Montréal, Montréal, Québec, Canada.

Email: jean-pierre.gage@umontreal.ca



concentration and attention than listening under ideal (quiet) environments (e.g., Desjardins & Doherty, 2013; Picou & Ricketts, 2014; Rakerd, Seitz, & Whearty, 1996; Xia, Nooraei, Kalluri, & Edwards, 2015). Measuring listening effort may be particularly well suited to comparing performance between two listening conditions when the speech recognition scores obtained on two tasks have reached ceiling levels. For example, the use of two hearing aids with different signal-processing algorithms may both yield the maximum correct recognition score on a given speech task administered in noise. However, the results of a behavioral listening effort task may reveal that performing the speech recognition task with one of the devices is more effortful than with the other one.

Presently, there is not a standardized test procedure to measure listening effort. Three broad categories of procedures have been used to measure listening effort: self-report, psychophysiological, and behavioral measures. A comprehensive review of these different approaches is beyond the scope of the present article but can be found in McGarrigle et al. (2014). In the present report, the analysis is limited to behavioral approaches. Measuring response times (RTs) is one approach that has been used to measure listening effort. However, this report will focus exclusively on the application of dual-task paradigms to measure listening effort during speech understanding. The decision to focus on this type of experimental paradigm was made because dual tasking (and even multitasking) is often required when processing speech in many real-life situations. This observation provides a form of ecological validity to the experimental procedure. Also, an informal literature search indicated that many investigators have used a dual-task paradigm to investigate listening effort.

The classic dual-task paradigm requires a participant to perform two tasks concurrently. One task is the primary task. In hearing research, this is usually the experimental listening task of interest (e.g., a task of speech recognition in noise). The other task, the secondary task, is used as a competing task. The tasks are administered to a participant under three experimental conditions: (a) the primary task is administered alone (primary-task baseline condition), (b) the secondary task is administered alone (secondary-task baseline condition), and (c) both the primary task and the secondary task are administered (dual-task condition). Typically, listening effort is calculated as the difference in performance on the secondary task between the baseline condition and the dual-task condition. Almost always, the listener is instructed to optimize performance on the primary task regardless of whether it is administered alone or under the dual-task condition. It is expected that performance on the primary task will be the same whether it is performed under the single-task condition or under the dual-task condition (Figure 1).

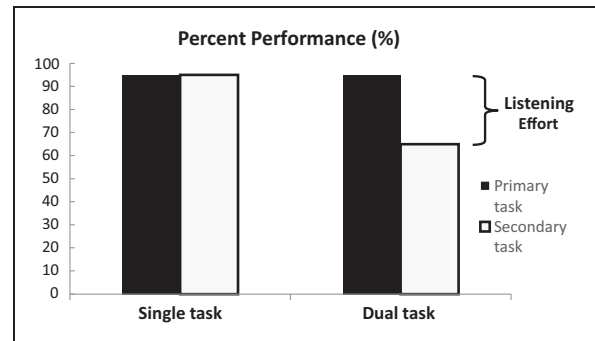


Figure 1. Illustration of the classic method used to measure listening effort.

The theoretical assumption underlying the use of a dual-task paradigm to measure listening effort is that the total processing resources that a person has available to perform tasks are limited in capacity and speed (Broadbent, 1958; Kahneman, 1973). If the attentional and the other cognitive resources required to perform the primary and the secondary tasks concurrently are less than the total resources available, then the person will be able to perform both tasks optimally. However, if the total resources required to perform both tasks exceed the maximum resources available, the person's processing system will prioritize one of the tasks under the dual-task testing condition. If instructed to optimize performance on the primary task, then a decrease in performance will be observed on the secondary task.

The focus of the present article is on the use of dual-task paradigms to measure listening effort during speech understanding. First, a summary is provided of the main findings reported by investigators that have employed this type of experimental paradigm to measure listening effort for speech among younger and older adults. Then, some unresolved issues related to the use of dual-task paradigms to measure listening effort during speech understanding are discussed.

Methods

In early 2015, a scoping search was undertaken to retrieve all studies that had used a dual-task paradigm to investigate listening effort for speech understanding. The review procedure did not strictly follow the guidelines for systematic reviews, such as the PRISMA guidelines (<http://www.prisma-statement.org/>). For example, no exact records were kept of the number of publications initially retrieved from each of the searched databases. Furthermore, an evaluation of the quality and strength of the reviewed studies was not undertaken. Nonetheless, a structured approach was applied to retrieving, including, and reviewing the publications. First, PubMed was queried for publications containing the keywords *effort*,

ease of listening, cognitive load, or processing load in combination with *hearing, listening, speech recognition, speech understanding, or speech perception*, and additionally in combination with *behavioral, dual task, response delay, or response time*, resulting in approximately 80 search results. Furthermore, Google scholar was queried with some of the above listed keywords to browse for publications not detected in the initial PubMed search. In total, approximately 90 publications were identified as potential articles for the review. From this initial list, studies were excluded if they were not peer reviewed, did not include the use of a dual-task paradigm to investigate listening effort, or if the primary task was not a speech-understanding task. Inclusion or exclusion of articles was based on information provided in the title and abstract. If these sources of information were inconclusive, the whole article was scanned for fulfilment of the aforementioned requirements. Each article was rated by one of the authors of the present review. In cases of uncertainty, consensus about the article's suitability was reached by all authors of the current article. Throughout the review process, the reference sections of the included articles were scanned for additional relevant articles.

Ultimately, 25 of the retrieved articles with adult participants were kept for the review. Furthermore, four articles that reported performance for adult participants and that were published during the review process were added to the list of documents reviewed. Hence, the present review of studies on adults is based on 29 articles (see Table 1). Publishing dates of the reviewed studies ranged from 1982 (Downs, 1982) to 2016 (Desjardins, 2016), and the majority of the studies were published after 2010. In addition to the dual-task studies conducted with adults, we retrieved another six articles reporting dual-task studies on listening effort for speech understanding in which the participants were children (Choi, Lotto, Lewis, Hoover, & Stelmachowicz, 2008; Hicks & Tharpe, 2002; Howard, Munro, & Plack, 2010; Hughes & Galvin, 2013; McFadden & Pittman, 2008; Stelmachowicz, Lewis, Choi, & Hoover, 2007). In the current review, detailed reports and discussions of the results will only consider studies in which the participants were adults.

The publications that had been judged to be eligible for the current review were read by at least one of the authors to extract the following information incorporated into a Master Table (see Table 1): participant characteristics (e.g., age, hearing status, other relevant factors), research question addressed in the study, primary task used, secondary task used, test conditions applied, other relevant dependent variables measured, main findings reported from the dual-task data, other related findings reported by the investigators, and any other relevant comments related to the article.

Results

The reviewed articles were evaluated regarding their findings on listening effort during speech understanding on the one hand and regarding methodological issues connected to dual-task paradigms on the other hand.

Summary of Findings

Effects of age. Several studies compared the amount of listening effort expended by younger and older adults when performing a speech-understanding task (Anderson-Gosselin & Gagné, 2011a, 2011b; Desjardins & Doherty, 2013; Helfer, Chevalier, & Freyman, 2010; Rakerd et al., 1996; Tun, McCoy, & Wingfield, 2009; Tun, Wingfield, & Stine, 1991). Generally, the results of these studies indicate that older adults (with normal or near-normal pure-tone detection thresholds) expend a greater amount of listening effort to recognize speech in noise than younger adults. This finding holds both for when the primary and the secondary tasks are administered under identical conditions (e.g., at the same signal-to-noise ratio [SNR]) to the two age-groups and when both groups are tested at the same level of performance by adjusting the SNR at which a given participant performs the speech task (Anderson-Gosselin & Gagné, 2011a, 2011b).

Effects of the hearing status. In some studies, a dual-task paradigm was used to investigate the effects of hearing status on listening effort among adult participants (Desjardins & Doherty, 2013; Helfer et al., 2010; Neher, Grimm, & Hohmann, 2014; Picou, Ricketts, & Hornsby, 2013; Tun et al., 2009; Xia et al., 2015). In general, the results of these investigations revealed that listeners with hearing loss deploy more listening effort than their age-matched counterparts with normal hearing acuity, especially when the speech-understanding task is administered in noise.

Effects of the perceptual modality in which the speech is presented. Some investigators compared the listening effort expended by participants as a function of whether the speech stimuli were presented audiovisually or in an auditory-alone modality (Anderson-Gosselin & Gagné, 2011a; Fraser, Gagné, Alepins, & Dubois, 2010; Picou, Ricketts, & Hornsby, 2011, 2013).

The results reported by Gagné and his collaborators revealed that when the speech stimuli were presented at the same SNR in both perceptual modalities, the listening effort expended to understand speech was greater in the auditory-alone modality (Anderson-Gosselin & Gagné, 2011a; Fraser et al., 2010). However, when both modalities were tested at the same primary-task performance level (i.e., poorer SNR in the audiovisual condition), speech understanding was more effortful in

Table 1. Summary of Published Articles in Which a Dual-Task Experimental Procedure Was Used to Measure the Listening Effort Expended to Perform a Speech-Understanding Task.

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Anderson-Gosselin & Gagné, 2011a) Effect of age (younger vs. older listeners with NH*)	Experiment 1: YNH; N = 25, 18–33 (23.5) years ONH; N = 25, 64–76 (69.0) years Experiment 2: YNH; N = 25, 20–43 (24.9) years ONH; N = 25, 65–77 (69.4) years	P: closed-set keyword recognition at SNR for 80% correct; Exp1: auditory-only, Exp2: AV; outcomes: pDTC for %-correct and RTs S: closed-set tactile pattern recognition; outcomes: pDTC for %-correct and RTs	Subjective scale (0%–100%) for amount of effort	Exp1 P: pDTC equal in YNH and ONH for %-correct and RTs S: ONH higher pDTC (greater effort) than YNH for %-correct; pDTCs equal in YNH and ONH for RTs Exp2 P: pDTC equal in YNH and ONH for %-correct and RTs S: pDTCs equal in YNH and ONH for %-correct; ONH higher pDTC (greater effort) than YNH for RTs	pDTCs generally higher for secondary (tactile) task than for primary task; pDTCs generally higher for AV conditions compared with auditory-only conditions
(Anderson-Gosselin & Gagné, 2011b) Effects of age (younger vs. older participants with NH); Equal SNR versus SNR that yields same performance level	YNH; N = 25, 18–33 (23.5) years ONH; N = 25, 64–76 (69.0) years	P: closed-set keyword recognition; cond1: equated level (–12 dB SNR), cond2: equated performance (SNR for 80% correct); outcomes: pDTC for %-correct and RTs S: closed-set tactile pattern recognition; outcomes: pDTC for %-correct and RTs	Subjective scale (0%–100%) for amount of effort	Cond1 P: ONH lower %-correct and longer RTs than YNH; pDTC: equal in YNH and ONH for %-correct and RTs S: ONH lower %-correct and longer RTs than YNH; pDTC: ONH higher pDTC (greater effort) for %-correct and RTs Cond2 P: ONH and YNH equal %-correct; ONH longer RTs than YNH; pDTC: equal in YNH and ONH for %-correct and RTs S: ONH lower %-correct and longer RTs than YNH; pDTC: ONH higher pDTC (greater effort) for %-correct, equal in YNH and ONH for RTs	No correlations between subjective ratings of effort and dual-task measures.
(Baer et al., 1993) Effects assessed: signal processing, SNR	OH; N = 5; 63–72 (67.2) years	P: 4-word sentences (auditory); closed-set word recognition, choose from visually displayed alternatives after S-task reaction; SNRs: 0, 3, 6, 9, 12 dB in each condition; 3 conds: control, spectrally enhanced (ENH), spectral enhancement with compression (ENHC) S: judge whether P-task sentence silly or sensible; outcome: absolute RTs for correctly judged, correctly identified sentences	None	P: %-correct identification higher for ENH and ENHC than for control, %-correct identification equal for ENH and ENHC S: RTs shorter for ENH and ENHC than for control; RTs shorter for ENHC than for ENH; RTs shorter for higher SNRs, however, also P-task performance increased with increasing SNR, thus effect of intelligibility and effort cannot be separated.	Benefit from ENH processing twice as large for RTs as for %-correct identification

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Desjardins, 2016) Effects assessed: directional microphone, noise reduction algorithm	OHI; N = 15; 54–76.8 (65.4) years	P: Harvard or IEEE sentences in 4-talker female babble S: visual motor tracking; outcome: %-time on moving target; effort: change in outcome from baseline to DT	Short Portable Mental Health Status Questionnaire (Pfeiffer, 1975). The digit symbol substitution test (DSST; Wechsler Adult Intelligence Scale-III; Wechsler, 1997) The letter number sequencing test (LNS) was used to measure participants' working memory performance. Self-reported estimates of listening effort expended on the speech-in-noise task were measured using a seven-category scaling procedure.	Directional microphone reduced listening effort there was a trend for participants to expend less listening effort with the NR algorithm activated in the hearing aids (24%—but not significant) Participants' listening effort was significantly reduced (by $\approx 5\%$) with the combined use of the directional microphones and the NR algorithm (NRDM) than with noise processing activated in the hearing aids (NoP). This is mostly accounted for by the DM effect	No correlation between rating scale and dual task cost However, participants did not perceive a difference in listening effort in background noise with and without the respective noise processing schemes (i.e., NR, DM, and NRDM) activated in the hearing aids. We did not find any statistically significant relationships between listening effort with NR, DM, or NRDM and cognitive function (processing speed and working memory). Experiences hearing aid users had worn hearing aids with directional microphones and NR algorithms, bilaterally, for at least 6 months fitted bilaterally with Starkey HA.
(Desjardins & Doherty, 2013) Effects assessed: masker type, low or high context, age group, hearing status	YNI; N = 15; 18–25 (21.7) years ONI; N = 15; 55–77 (66.9) years OHI; N = 15; 59–76 (68.2) years, experienced HA users, bilaterally fitted with own devices	P: R-SPIN at SNR for 76% correct; maskers: 6-talker babble, 2-talker babble (TTB), SSN; outcome: %-correct target words S: visual motor tracking; outcome: %-time on moving target; effort: change in outcome from baseline to DT	Subjective scale (0–100) for “ease of listening”; RSPAN for WM function; Digit-Symbol Substitution Test (DSST) for perceptual speed of processing; Visual STROOP test for selective attention	P: constant performance across conditions and groups; performance better for high-context SPIN sentences in all groups S: ONH and OHI higher effort than YNH in TTB and SSN maskers; equal effort for low- and high-context R-SPIN sentences in all groups; ONH and OHI higher effort in SSN and TTB than in six-talker babble; YNH higher effort in six-talker babble than in SSN; effort equal for SSN and TTB	SSN: effort-RSPAN ($r = -.30^*$); DPRT-DSST ($r = -.33^*$); TTB: effort-RSPAN ($r = -.25^*$); Subjective ratings: SSN easiest, TT hardest; All results for complete study sample;
(Desjardins & Doherty, 2014) Effects assessed: signal processing, low or high context, performance level	OHI; N = 12; 50–74 (66.0) years, experienced HA users, bilaterally fitted with own devices	P: R-SPIN at SNR for 76% and 50% correct; with or without NR; masker: 2-talker babble (TTB); outcome: %-correct target words S: visual motor tracking; outcome: %-time on moving target; effort: change in outcome from baseline to DT	Subjective scale (0–100) for “ease of listening”; RSPAN for WM function; Digit-Symbol Substitution Test (DSST) for perceptual speed of processing	P: equal performance with or without NR; word recognition higher for high-context R-SPIN sentences S: without NR: higher effort in 50%-correct than in 76%-correct condition; with NR: no difference in effort between conditions; at $\approx 76\%$ correct: lower effort with NR; at 50%-correct: equal effort with/without NR; equal effort for low- and high-context R-SPIN sentences	Effort: at 50%-correct with NR-DSST ($r = -.58^*$), not significant when correcting for multiple comparisons; Subjective ratings: 76% correct condition easier than 50% correct; no difference with/without NR

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Downs, 1982) Effects assessed: spatial configuration, low or high context assessed: amplification	HI; N = 23; 29–68 (51.0) years, experienced HA users, bilaterally fitted with study devices	P: Monosyllabic words; masker: 8-talker babble; targets: 45 dB HL; SNR: 0 dB; outcome: % correct S: RTs to visual probe (flashing light); effort: change in outcome from baseline to DT	None	P: better word recognition with HA than without ($p < .001$) S: lower effort with HA than without ($p < .05$), 16% of variance in decreased RT due to HA use; however, note that word recognition with versus without HA was not at the same performance level	None
(Feuerstein, 1992), Effects assessed: spatial configuration, low or high context	YNIH; N = 48, 18–24 (19) years	P: R-SPIN sentences; –5 dB SNR; dichotic: target at 65° to left or right, masker at 65° to other side; conditions: binaural (BIN), monaural ^a near (MN) or far (MF), unoccluded ear closer to target speaker in MN, closer to masker speaker in MF; outcome: %-correct target words S: push button to turn off light once turned on; outcome: RTs; “attentional effort” = shift in RT from baseline to dual task	Perceived ease of listening ratings; rating scale from 0 (very, very difficult) to 100 (very, very easy); modified direct magnitude estimation (MME)	P: performance best for BIN, poorest for MF ($p < .01$); better performance for high context than for low-context sentences; interaction of context- and listening conditions S: attentional effort higher in MF than in BIN and MN; attentional effort in BIN and MN equal (all $p < .01$);	BIN rated as easiest, MF rated as most difficult ($p < .01$); Rated ease-attentional effort: $r = -.01$, controlling for recognition performance; attentional effort – word recognition: $r = -.19$, controlling for perceived ease; word recognition-perceived ease: $r = .71$, controlling for attentional effort
(Fraser et al., 2010) Effects assessed: modality (SNR vs. performance)	Experiment 1 NH; N = 30, 18–41 (25.0) years Experiment 2 NH; N = 30, 18–45 (25.0) years	P: closed-set keyword recognition; Expl: equated level (–11 dB SNR) for A-only (80% correct) and AV (96% correct); Exp2: equated performance (80%) at –11 dB SNR for A-only and –19 dB SNR for AV; outcomes: %-correct and RTs S: closed-set tactile pattern recognition; outcomes: %-correct and RTs	Subjective scale (0%–100%) for amount of effort	Expl 1 P: S: decrease in %-correct and increase in RTs by dual-task equal for A-only and AV Expl2 P: decrease in %-correct by dual task for AV but not for A-only; increase in RTs by dual-task equal for A-only and AV S: decrease in %-correct and increase in RTs by dual task bigger in AV than in A-only	Expl P: overall, %-correct higher in AV than in A-only; RTs equal for AV and A-only S: overall, %-correct and RTs equal for AV and A-only Expl2 P: overall, %-correct equal for AV and A-only; RTs longer for AV than for A-only S: overall, %-correct lower for AV than for A-only; RTs longer for AV than for A-only

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Helfer et al., 2010) Effects assessed: age group by hearing status, masker characteristics, spatial configuration	YNH; N = 10, 20–38 (23) years ONH (some high-frequency mild losses); N = 10, 60–69 (63) years	Setup: TVM sentences; free field; 3 talkers=1 target, 2 distractors; conds: all talkers at front (FF) or target at front, distractors at 60° right front (F-RF) P1: indicate number (0–2) of distractor sentences played backwards; outcome: % correct P2: repeat sentence of designated target talker P1 and P2 performed separately (single-task) or congruently (dual-task)	Participants instructed to divide their attention equally between both tasks	P1: YNH performed better than ONH; F-RF performance better than FF; single-task better than dual-task; P2: YNH and ONH performance equally, F-RF performance better than FF, but only for single-task; single-task better than dual-task; basically no spatial-separation benefit for ONH in dual-task Trend for higher dual-task cost in YNH than in ONH and for F-RF compared with FF condition; Spatial-separation benefit reduced by dual-task; trend for ONH to have greater spatial-separation benefit than YNH in P1 but smaller in P2	Age associated with dual-task cost ($i = .90^{***}$) in P1 FF condition, that is, the older, the higher the cost of dividing attention For P2, ONH had a much greater disadvantage when the distractor sentences were played forward (intelligible) than when played backwards, especially in FF cond, thus much bigger susceptibility to semantic interference
(Hornsby, 2013) Effects assessed: amplification, signal processing, secondary task	HI; N = 16, 47–69 (65) years; fit with Phonak Micro Exelia BTE HAs;	P: word recognition; 8–12 words per list; target speech at 0°, cafeteria babble at 60°, 120°, 180°, 240°, 300°; 3 conds: without HA, with basic HA setting, with advanced HA setting; SNR for 70% correct in advanced HA setting SI: reaction to visual probe; outcome: % change from single- to dual-task RTs S2: after SI, recall 5 last words of P: task from memory	Subjective rating before and after testing; 5 items for concentration, listening effort, distractibility, ability to maintain focus, current state of mental draining	P: performance better in aided than unaided conditions; no difference between aided conditions; SI: dual-task cost in RTs was higher in unaided and basic HA conds than with advanced HA settings S2: word recall better with than without HAs; no difference between HA settings	P: performance increased over the course of testing SI: dual-task cost in RTs increased over the course of testing in the unaided but not in the aided conditions S2: word recall remained stable over testing blocks Large variance in SI effects across participants Increase in subjective fatigue and difficulty to maintain attention post- compared to pre-dual-task testing
(Neher et al., 2014) ^c Effects assessed: hearing status, cognitive status, signal processing, SNR	OHI H+C+; N = 10, 60–79 (72.1) years OHI H-C+; N = 10, 70–83 (74.7) years OHI H+C-; N = 10, 68–81 (75.8) years OHI H-C-; N = 10, 65–81 (75.0) years	P: OLSA matrix test sentence recognition at -4, 0, 4 dB SNR; each SNR with inactive or moderate or strong NR task; visual response time (VRT) respond differently to even or odd digits; outcome: absolute RTs	Reading span, used for group assignment (cognitive function), not as an outcome measure; Paired-comparisons of preference ratings for NR settings	P: worse performance ($p < .0001$) with strong NR than with inactive NR; performance equal for moderate NR and inactive NR; performance increased with increasing SNR S: VRTs decreased with increasing SNR, differences between all SNRs ($p < .001$); VRTs longer for strong NR than for inactive NR ($p < .001$); VRTs equal for inactive NR and moderate NR	P: H+C+ group performed better than H-C+ ($p < .05$) and H-C- ($p < .001$) groups; thus, differences driven by HL rather than cognition; Some NR preferred over no NR for all SNRs and all groups, strength of preferred NR dependent on SNR and group

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Neher, Grimm, et al., 2014) ³ Effects assessed: hearing status, cognitive status, signal processing, SNR	OHI H+C+; N = 10, 60–80 (73.0) years OHI H-C+; N = 10, 71–84 (75.0) years OHI H+C-; N = 10, 69–82 (76.6) years OHI H-C-; N = 10, 66–82 (75.5) years	P: OLSA matrix test sentence recognition at -4 and 0 dB SNR; each SNR in 5 NR conditions with processing in signal and noise S: visual response time (VRT) task; visually displayed digits; respond differently to even or odd digits; outcome: RTs relative to median of several baseline conditions	Subjective effort ratings for sentences at -4, 0, 4 dB; 9-point rating scale ranging from "completely effortless" to "maximally effortful"; Reading span, used for group assignment (cognitive function), not as an outcome measure; Paired-comparisons preference ratings for NR settings	P: performance different for different NR settings, best for NR processing on noise only (no processing on signal); performance better (p < .0001) at 0 dB than at -4 dB SNR; S: VRTs shorter (p < .05) at 0 dB than at -4 dB SNR; VRTs longer (p < .01) with NR processing in signal and noise than for other conditions; other conditions equal to each other	P: performance of H+C+ group better than of H-C+ (p < 0.01) and H-C- (p < 0.001) groups. S: VRTs equal for listener groups with increasing SNR; effort ratings different for different NR settings, lowest for NR processing in signal and noise; no differences between groups NR processing in signal and noise preferred over all other NR settings; no differences by group or SNR
(Ng et al., 2015) Effects assessed: signal processing (NR), masker language, WM span (high or low)	OHI; N = 26, 56–65 (62.4) years; experienced HA users	P: repeat sentence-final word (for half of the sentence lists only); conds: 4-talker babble (language same as target speech or Cantonese), with or without binary masking NR; at individual SNR for 95% word recognition; outcome: %-correct word recognition S: free recall of all sentence-final words from last set of P-task sentences; outcome: %-correct	Reading span for WM capacity, division into low-span and high-span group; Note: only for half the sentence lists there was an actual dual-task situation. For the other half, there was only the S-task	P: performance with same-language babble without NR processing poorer (ca 93% correct) than for other conds. (close to 100% correct) S: in same-language babble, recall better with than without NR; in Cantonese babble, no difference between with or without NR; high-span group outperformed low-span for primacy list items, but not for asymptote; for low-span (but not for high-span) group performance better with than without NR for recency items, but not for primacy or asymptote;	Different strategies in recall by span group: low-span group tendency to recall words from late list positions, no such effect in high-span group
(Pals et al., 2013) Effects assessed: signal processing, secondary task	YNH; N = 19, 19–25 (22) years CI simulation (noise vocoding)	P: Recognition of vocoded sentences; conds: 2, 4, 6, 8, 12, 16, or 24 channels or no processing (control); at self-adjusted level 65–75 dB SPL; outcome: %-correct words S1: rhyme judgments for visually presented monosyllabic words; button press yes or no; outcome: RTs for correct responses S2: mental rotation of Japanese characters; judgment whether two displayed characters are the same when rotated; button press yes/no; outcome: RTs for correct responses	Rating of perceived workload; NASA task-load index (TLX)	P: no difference in performance by choice of secondary task; performance improved from 2 to 4 channels and from 4 to 6 channels, for 6 or more channels stable at ceiling S: RTs decreased from 2 to 4, 4 to 6, and 6 to 8 channels; thus, RTs decreased where speech recognition stable; reduction of RTs by increased no. of channels bigger for RTs during primary task than between primary-task trials	S: Training effects in RTs for both secondary tasks over the course of the experiment Both dual-task conditions judged as more effortful than the single-task conditions; ratings equal for the two secondary tasks; decrease in effort ratings for conditions up to 6 channels

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Pals, Sarampalis, van Rijn, & Baskent, 2015) Effects assessed: performance level, noise type, # of vocoded channels for speech recognition task	YNH; N = 19, M = 19 (18–25) years	P: Sentence understanding; conds: quiet, SSN, 8-talker babble, at individual SNRs for 79% correct and for near ceiling (NC) S: rhyme judgments for pairs of visually presented monosyllabic words; outcome: RTs for correct responses, separate for judgments during vs. between P-task stimuli	WAIS for processing speed; reading span for working memory capacity Note: this study is also listed under the single-task studies for response delays	P: intelligibility as intended S: RTs longer in masker than in quiet; no effect of noise type or intelligibility level on RTs	WAIS predicted RTs
(Picou et al., 2011) Effects assessed: modality, masker type, effect of WMC and lip-reading ability on A-only versus AV effort	NH; N = 20, 19–44 (27.9) years	P: recognition of monosyllabic words in quiet or 4-talker babble; sets of 5 words; individual SNR for 75% correct; conds: A-only and AV; outcome: %-correct S: Recall of presented words (sets of 5 words); outcome: %-correct	Rating of degree of effort put into hearing what was said; 11-point scale from “no effort” to “lots of effort”; Lipreading skills, Revised Shortened Utley Sentence Lipreading Test (ReSULT), visual-only sentence recognition WM capacity, automated operation span task (AOSPAN), solving of simple equations, memorization of letters	P: performance better in quiet than in babble; performance equal for A-only and AV conditions; no effect of serial position within set of words S: word recall better in quiet than in noise; effect of serial position; better recall for positions 4 and 5 than for positions 1–3; recall equal for A-only and AV conditions	Subjectively rated effort larger in babble than in quiet; no difference in effort ratings for A-only versus AV testing or by serial position; P: In babble: those with better lipreading benefited more from provision of visual cues (AV compared to A-only) in primary task S: deviation into low-AOSPAN and high-AOSPAN groups: in babble, high-AOSPAN group had higher recall benefit from provision of visual cues (AV compared to A-only), but in quiet, recall for high-AOSPAN group worse in AV than in A-only
(Picou et al., 2013) Effects assessed: modality, amplification	HI; N = 27, 49–89 (65.3) years Fitted with Phonak Savia 211 BTEs	P: recognition of monosyllabic words in quiet and 4-talker babble; individual SNR for 50%–70% correct; A-only and AV; aided and unaided testing S: reaction to visual probe: react to red rectangle but not to white; outcome: RTs to probe trials	Lipreading skills, ReSULT, (see previous entry) WM capacity, AOSPAN (see previous entry) Verbal processing speed, lexical decision task (LDT), real word or not? outcome: RTs	P: performance better in quiet than in babble, effect bigger for AV than A-only condition; in quiet, AV performance better than A-only, but not in babble; HA benefit larger in A-only than in AV condition, but small in both; HA-benefit benefit larger in quiet than in babble S: RTs equal for A-only and AV conditions; RTs shorter in quiet than in babble; RTs shorter with than without HAs Overall, RT effects small and with big variance across listeners	S: HA benefit in RTs for A-only and AV correlated with each other; HA benefit in RTs associated with processing speed; for unaided quiet, RTs of listeners with better lipreading and faster verbal processing speed benefited more from visual cues; for unaided quiet and aided babble, RTs of listeners with smaller AOSPAN benefited more from visual cues

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Picou & Ricketts, 2014) Effect of secondary- task complexity on paradigm sensitivity to listening effort. Effects assessed: modality, hearing status, secondary task	YNH; N = 17, 21–24 (23.0) years HI; N = 17, 23–73 (60.1) years; sensorineur at HL, unaided testing	P: recognition of monosyllabic words in quiet or 4-talker babble; individual SNR for 80% correct; A-only and AV; outcome: %-correct S1 (simple): reaction to visual probe; react to red rectangle but not to white S2 (complex): reaction to visual stimulus; choice of response button based on even or odd digit S3 (semantic): semantic-category judgment of primary-task word, hit response button if word was a noun Outcome for all S-tasks: RTs	None	YNH P: performance better in quiet than in babble; performance equal for all S-task conditions; performance equal for A-only and AV conditions S: RTs longer for semantic than for complex and simple S-tasks and also longer for complex than for simple; RTs equal in quiet and babble for simple and complex S-tasks, but longer in babble than in noise for semantic S-task; RTs equal for A-only and AV in all S-tasks HI P: performance better in quiet than in babble; in quiet, performance better for AV than A-only, but equal in babble; performance equal for all S-task conditions S: RTs longer in babble than in quiet for all S-tasks; RTs longest for semantic S-task, intermediate for complex S-task, shortest for simple S-task; RTs equal in A-only and AV conds for all S-tasks	None
(Picou, Gordon, & Ricketts, 2016) Effect of noise and reverberation	YNH; N = 18, 22–30 (24.8) years	P: recognition of monosyllabic words, outcome: % correct; in quiet, 3 levels of reverberation (low, moderate, high); one SNR with low reverberation, two SNRs with moderate reverberation, two SNRs with high reverberation, see RQs; S: word-class judgment of primary-task word, hit response button if word was a noun; outcome: RTs (all button presses, not only correct)	RQ A: effect of babble (vs. quiet) with low or moderate rate or high reverberation at constant intelligibility (ca 84%) RQ B: for constant moderate reverberation, effect of babble (84% or 77% correct) vs. quiet; RQ C: at constant SNR, effect of low vs. moderate reverberation; RQ D: at constant SNR, effect of moderate vs high reverberation	P: not constant 84% as intended across reverberation levels (RQ A), poorer performance with high reverberation S: RQ A: RTs longer in babble than in quiet for all reverberation levels; no difference between RTs for different reverberation levels, neither in quiet nor babble RQ B: for constant reverberation, longer with babble than in quiet, longer for 77% correct than for 84% correct RQ C and D: for constant SNRs, no difference in RTs between reverberation levels	RTs not correlated with word recognition, neither in quiet nor in babble and at no reverberation level; RTs correlated with each other, across babble conditions and reverberation levels; Word recognition scores not correlated with each other across conditions; Age not correlated with RTs or word recognition scores (note, all YNHI)

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Rakerd et al., 1996) Effects assessed: Exp1: hearing status, Exp2: hearing status by age	Exp1 YNH; N = 8, ages not reported YHI; N = 9, ages not reported Exp2 YNH; N = 11, 21–29 (24) years HI; N = 11, 52–73 (62) years, presbycusis, mostly HA users	P: conds: listening 60 sec to (1) steady-state speech noise, (2) speech passages, questions on passages answered after digit recall; stimuli at 65 dB SPL for YNH, at most comfortable loudness for YHI and HI S: digit recall; string of digits presented before P-task, ordered recall after P-task; Exp1: 9-digit strings, Exp2: 9–13-digit strings; outcome: number correct recalls	Listeners instructed to give equal priority to speech listening and digit recall	Exp1 S: digit recall better after noise than after speech trials; digit recall equal for YNH and YHI after noise trials, poorer for YHI than for YNH after speech trials; difference between noise and speech trials bigger for YHI than for YNH Exp2 S: recall better for YNH than for OHI; digit recall better after noise than after speech trials, effect bigger for OHI than for YNH	None
(Rigo, 1986) Compare speechreading under focused attention and divided attention	Participants = 30 normal-hearing young adults	Stimuli were low-pass filtered 22 consonants presented in a C-v context 4 conditions: AV, V, A, V with concurrent auditory processing PRIMARY task was an auditory syllable detection task V-AP condition required the performance of the VCV lipreading task with attention divided between auditory and visual modalities. Subjects were provided a second set of written instructions prior to administration of the end detection practice list and V-AP test condition. Subjects were instructed to maintain the same level of accuracy on the end detection task under the V-AP condition as that achieved on the practice list.	none	Results indicate that lipreading performance during divided attention was significantly lower than that measured during focused attention. The performance decrement suggests that simultaneous processing of the lipreading and auditory tasks exceeded capacity; with attention divided between modalities a sufficient amount of capacity was not available for optimal processing of visual stimuli, causing lipreading performance to suffer.	(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Sarampalis et al., 2009) Effects assessed: signal processing, low or high context, SNR, secondary task (?)	Exp 1 YNIH; = 25, 18–26 (20.0) years Exp 2 YNIH; = 25, 19–27 (ca. 21) years	Exp 1 P: R-SPIN sentences; conds: in quiet at 65 dB SPL, in 4-talker babble at $-2/+2$ dB SNR, with babble at 65 dB SPL; in babble unprocessed or with NR; outcome: %-correct targets S: remember last word of each sentence; recall after 8 sentences; outcome: %-correct Exp 2 P: IEEE sentences; conds: in quiet at 65 dB SPL, in 4-talker babble at $-6, -2, \text{ or } +2$ dB SNR, with babble at 65 dB SPL; in babble unprocessed or with NR; outcome: %-correct targets S: complex visual RT task for speed of processing; different response for odd/even digit; arrow pointing toward/away from digit; outcome: RTs	None	Exp 1 P: performance better at high SNR than at low SNR; performance perfect for low- and high-context sentences in quiet; in babble better for high- than for low- context; negative effect of NR at -2 dB SNR independent of context S: recall better for high- than for low- context words; for low- context: recall better in quiet than in babble, better at 2 dB SNR than at -2 dB SNR, almost no effect of NR; for high-context: recall better in quiet than in babble, with NR recall equal at -2 and $+2$ dB SNR, without NR, recall better at $+2$ dB SNR, recall equal with and without NR at 2 dB SNR, but recall poorer with NR at -2 dB SNR Exp 2 P: performance perfect in quiet; performance decreased with decreasing SNR; no effect of NR S: RTs increased with decreasing SNR; effect of NR only at lowest SNR (-6 dB SNR); RTs shorter with NR	None
(Seeman & Sims, 2015) Comparison of psychophysiological and dual-task on listening effort	YNIH; N = 46 (divided into 3 groups), 18–38 (21.2) years	Group 3: DT experiment (N = 15 or 16) P: sentences-in-noise (SIN) at $+15, +5$ dB SNR, in steady-state noise; outcome: total number of correct keywords (max 75, 5 per sentence) S: visual letter identification; press button when displayed letter is target letter; outcome: %-correct and RTs, RT outcomes proportional: (RT _{dual} – RT _{single})/RT _{single}	Group 1: diotic-dichotic digit listening with psychophys measures Group 2: SIN understanding with psychophys measures Psychophys measures: skin conductance (SC), heart-rate variability (HRV), heart rate (HR) Subjective: NASA Task Load Index (TLX), except physical demand scale None	P: 100% correct at $+15$ dB SNR, 83% correct at $+5$ dB SNR S: at $+15$ dB SNR RTs not different from baseline; at $+5$ dB SNR, RTs longer than at baseline and at $+15$ dB SNRExp 1 P: performance better at high SNR than at low SNR; performance perfect for low- and high-context sentences in quiet; in babble better for high- than for low- context; negative effect of NR at -2 dB SNR independent of context	NASA TLX: higher load for dual than for single tasks, load increased for decreasing SNRs (Groups 2 and 3), load increased with task complexity for dichotic digits (Group 1) HRV was sensitive to dichotic task complexity (Group 1) and to high ($+15/+10$ dB) vs low ($+5/0$ dB) SNRs (Group 2); HR and SC had some sensitivity No correlations of psychophys measures or RTs with NASA TLX None

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Tun et al., 1991) Effects assessed: age, proposition density in continuous passages	YNH; N = 18, 18–20 (18.3) years ONH; N = 18, 60–80 (69.6) years	<p>Exp 1 P: R-SPIN sentences; conds: in quiet at 65 dB SPL, in 4-talker babble at -2/+2 dB SNR with babble at 65 dB SPL; in babble unprocessed or with NR; outcome: %correct targets</p> <p>S: remember last word of each sentence; recall after 8 sentences; outcome: %correct</p> <p>Exp2 P: IEEE sentences; conds: in quiet at 65 dB SPL, in 4-talker babble at -6, -2, or +2 dB SNR with babble at 65 dB SPL; in babble unprocessed or with NR; outcome: %correct targets</p> <p>S: complex visual RT task for speed of processing; different response for odd/even digit; arrow pointing toward/away from digit; outcome: RTs</p>	<p>Reading and listening span for working memory Group 1: diotic-dichotic digit listening with psychophys measures</p> <p>Group 2: SIN understanding with psychophys measures</p> <p>Psychophys measures: skin conductance (SC), heart-rate variability (HRV), heart rate (HR)</p> <p>Subjective: NASA Task Load Index (TLX), except physical demand scale</p>	<p>S: recall better for high- than for low- context words; for low- than in babble, better at 2 dB SNR than at -2 dB SNR, almost no effect of NR; for high-context: recall better in quiet than in babble, with NR recall equal at -2 and +2 dB better at +2 dB SNR, recall equal with and without NR at 2 dB SNR, but recall poorer with NR at -2 dB SNR</p> <p>Exp2 P: performance perfect in quiet; performance decreased with decreasing SNR; no effect of NR</p> <p>S: RTs increased with decreasing SNR; effect of NR only at lowest SNR (-6 dB SNR): RTs shorter with NR</p>	<p>Reading and listening span combined into one WM span measure; YNH had larger WM spans than ONH; WM span associated with proposition recall in single-task ($r = .55^{**}$) and dual-task ($r = .33^{**}$); WM span associated with simple-RT performance in dual-task ($r = -.40^{**}$) and with choice-RT performance in single-task ($r = -.52^{**}$) and dual-task ($r = -.54^{**}$); NASA TLX: higher load for dual than for single tasks; load increased for decreasing SNRs (Groups 2 and 3), load increased with task complexity for dichotic digits (Group 1)</p> <p>HRV was sensitive to dichotic task complexity (Group 1) and to high (+15/+10 dB) vs low (-5/0 dB) SNRs (Group 2); HR and SC had some sensitivity</p> <p>No correlations of psychophys measures or RTs with NASA TLX</p>
		<p>S1: simple RT, press button, when letter "j" on screen</p> <p>S2: choice RT, press one button for letter "j", press different button for letter "H" Group 3: DT experiment (N = 15 or 16)</p> <p>P: sentences-in-noise (SIN) at +15, +5 dB SNR, in steady-state noise; outcome: total number of correct keywords (max 75, 5 per sentence)</p> <p>S: visual letter identification; press button when displayed letter is target letter; outcome: %correct and RTs, RT outcomes proportional: (RT_{dual} - RT_{single})/RT_{single}</p>	<p>Group 2: SIN understanding with psychophys measures</p> <p>Psychophys measures: skin conductance (SC), heart-rate variability (HRV), heart rate (HR)</p> <p>Subjective: NASA Task Load Index (TLX), except physical demand scale</p>	<p>P: recall better for low-density than for high-density passages, recall equal for both S-tasks; performance in all conditions similar for YNH and ONH but high-level propositions (in hierarchy) better recalled than low-level and this effect larger in ONH than in YNH; S: YNH faster than ONH; responses faster for simple-RT task than for choice-RT task; responses faster for low-density than for high-density passages but only for choice-RT task; dual-task cost was greater for ONH than for YNH for simple- RT task but not for choice-RT task; 100% correct at +15 dB SNR, 83% correct at +5 dB SNR</p> <p>S: at +15 dB SNR RTs not different from baseline; at +5 dB SNR, RTs longer than at baseline and at +15 dB SNR</p>	<p>Reading and listening span combined into one WM span measure; YNH had larger WM spans than ONH; WM span associated with proposition recall in single-task ($r = .55^{**}$) and dual-task ($r = .33^{**}$); WM span associated with simple-RT performance in dual-task ($r = -.40^{**}$) and with choice-RT performance in single-task ($r = -.52^{**}$) and dual-task ($r = -.54^{**}$); NASA TLX: higher load for dual than for single tasks; load increased for decreasing SNRs (Groups 2 and 3), load increased with task complexity for dichotic digits (Group 1)</p> <p>HRV was sensitive to dichotic task complexity (Group 1) and to high (+15/+10 dB) vs low (-5/0 dB) SNRs (Group 2); HR and SC had some sensitivity</p> <p>No correlations of psychophys measures or RTs with NASA TLX</p>

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Tun et al., 2009) Effects assessed: age group, hearing status, semantic relatedness of words in word lists	Y: N = 24, 20–46 (27.9) years (YNH, N = 12; YHI (mild); N = 12) O: N = 24, 67–80 (73.9) years (ONH, N = 12; OHI (mild), N = 12)	P: word list at 70 dB SPL, followed by 30-second counting task, followed by word recall; conds: semantically related and unrelated word lists; outcome: number of correctly recalled words S: visual motor tracking during responses to P-task (after counting task); outcome: % time on moving target; speed individually set to 50 - 60% tracking accuracy; effort: change in outcome from baseline to DT	NH and HI groups matched on 3 cognitive measures (backward digit span for verbal WM, word-list recall for episodic memory, trail making A and B for executive control)	P: recall better for related than for unrelated word lists; recall better in Y than in O listeners but effect of dual-task similar in both groups; age x list-type interaction: O greater benefit of word relatedness than Y; NH groups performed better than HI, effect bigger in Y than in O S: effort larger for O than for Y, i.e., larger dual-task cost in O; HI poorer performance than NH, especially in O group;	None
(Wild et al., 2012) ^d Effects assessed: secondary-task modality, signal processing (not in Has)	YNH; N = 21, 19–27 (21.0) years	P: sentence recognition; 4 vocoding conds: clear speech, six-band (NV-hi), six-band compressed (NV-lo), spectrally rotated (rNV); yes/no response for sentence intelligibility S1: auditory distracters: 400-ms noise bursts; non-targets: long onset, sharp offset; targets: sharp onset, long offset; diotic to P-task stimuli; yes/no response whether a target in trial S2: visual distracters: 200-ms presentations of cross-hatched white ellipses on black background; non-targets: solid lines; targets: dashed lines; yes/no response whether a target in trial P: listen to expository passages (ca. 230 words), repeat content later on in own words; two levels of proposition density S1: simple RT, press button, when letter ‘j’ on screen S2: choice RT, press one button for letter ‘j’, press different button for letter ‘H’	For each trial, visual prompt cued attention to a single stimulus stream: “Speech” for speech stimuli, “Chirps” for auditory distracters, “Footballs” for visual distracters. Next to behavioral results, fMRI data were gathered. Reading and listening span for working memory	Results reported for attended conditions P: clear speech and NV-hi equally intelligible, better than NV-lo, which was better than rNV S1 and S2: Performance better than chance; performance better for visual distracters than for auditory distracters; performance unaffected by condition in (unattended) P-task P: recall better for low-density than for high-density passages, recall equal for both S-tasks; performance in all conditions similar for YNH and ONH but high-level propositions (in hierarchy) better recalled than low-level and this effect larger in ONH than in YNH; S: YNH faster than ONH; responses faster for simple-RT task than for choice-RT task; responses faster for low-density than for high-density passages but only for choice-RT task; dual-task cost was greater for ONH than for YNH for simple-RT task but not for choice-RT task	None Reading and listening span combined into one WM span measure; YNH had larger WM spans than ONH; WM span associated with proposition recall in single-task ($r = .55^{**}$) and dual-task ($r = .33^{**}$); WM span associated with simple-RT performance in dual-task ($r = -.40^{**}$) and with choice-RT performance in single-task ($r = -.52^{**}$) and dual-task ($r = -.54^{**}$)

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
(Wu et al., 2014) Effects assessed: amplification by signal processing	Exp 1 OHI; N = 29, 56–85 (72.7) years; 25 were experience HA users Exp 2 OHI; N = 19, 56–85 (71.7) years; subset of Exp 1 participants Exp 3 YNIH; N = 14, 20–37 (23.4) years	Exp 1 P: Connected speech test (CST); % correct keyword repetition; stimuli prerecorded in car with noise at 75 dBA, SNR –1 dB, 3 conds: unaided, aided omnidirectional (OMNI), aided directional to back (DIR); testing with individually amplified speech (NAL- NLI); S: driving performance: mean, SD, interquartile range (IQR) of driving distance to lead vehicle; smaller numbers = better performance Exp 2 (as Sarampalis et al., 2009) P: Speech recognition; 3 conds: unaided, OMNI, DIR S: complex visual RT task; different response for odd/even digit; outcome: RT's Exp 3 as Exp 2	Participants were asked to pay equal amounts of attention to P and S tasks.	Exp 1 P: single-task performance better than dual-task; OMNI better than unaided, DIR better than unaided and OMNI S: baseline better than dual-task; dual-task performance and dual-task cost equal for all HA conditions Exp 2 P: performance equal for single- and dual- task; performance better for OMNI than for unaided and better for DIR than for OMNI and unaided S: baseline better than dual-task; dual-task performance and dual-task cost equal for all HA conditions Exp 3 P: performance better in DIR than in unaided and OMNI S: baseline better than dual-task; dual-task performance better (and cost lower) in DIR than in unaided and OMNI	RTs in Exp 2 were associated with driving performance in Exp 1 ($r = 0.5^*$) Benefit of DIR for RTs larger for YNIH in Exp 3 than for OHI in Exp 2
(Xia et al., 2015) Effects assessed: hearing status, spatial configuration, masker characteristics	NH; N = 8, 51–61 (Med = 55.0) years HI; N = 8, 59–66 (Med = 64.0) years +	P: CRM at 65 dB SPL (0 dB SNR), with NAL-R gain for HI listeners; masker: 2 competing talkers; 4 conds for target- masker relationship: (1) sex differences and spatial separation, (2) sex difference only, (3) spatial separation only, (4) no cues; spatial separation 15° in Exp 1, 60° in Exp 2; outcome: %-correct target words S: visual motor tracking; movement speed set to individual 60% accuracy; outcome: % time on moving target; effort: change in outcome from baseline to DT	None	Exp 1 (15° spatial separation) P: HI performed worse than NH in all conds but the no-cue cond ($p < .001$); in both groups, equal performance for both-cues and sex-cues conds; in HI, equal performance also for spatial-cues and no-cues conds; pattern true for both single-task and dual-task conds S: tracking-accuracy lower in dual-task cond; group – cue-type interaction: During listening of P-task: in NH, DT cost lower in both-cues than in spatial-cue cond, lower in no-cue than in spatial-cues and sex-cues conds; in HI, lower DT cost in both-cues than in sex-cues cond, lower DT cost in no-cue than in sex-cues and spatial-cues conds;	Exp 1 Spatial-separation benefit in P-task associated with reductions of cost in S-task ($r = .59^*$)

(continued)

Table 1. Continued

(Authors) Research question	Participants	Experimental tasks used	Other relevant measures or information	Findings for dual task	Other relevant findings
				<p>During response of P-task: same pat-term as during listening</p> <p>Exp2 (60° spatial separation)</p> <p>P: all conds different from each other, except for sex-cues and spatial-cues conds;</p> <p>S: During listening of P-task: DT-cost lower for both-cues than for sex-cues cond, lower for no-cue than for spatial-cues and sex-cues conds;</p> <p>During response of P-task: DT-cost equal for spatial-cue and sex-cue conds, lower for no-cue than for spatial-cues and sex-cues conds, lower for both-cues than for sex-cues conds</p>	

Note. AOSPAN = automatic operation span task; AV = audiovisual; BIN = binaural; BTE = behind the ear; C-/+/+ = cognitive function good (+) or cognitive function poor (-); CST = continuous speech test; DIR = directional; DM = directional microphone; DPRT = Digital pursuit Rotor Tracking; DT = dual task; DSST = digit symbol substitution test; ENH = spectrally enhanced; ENHC = spectrally enhanced with compression; FF = talker at front; F-RF = talker at right-front; H-/+/+ = hearing sensitivity good (+) or hearing sensitivity poor (-); HA = hearing aid; HI = hearing impaired; HRV = heart rate variability; LDT = lexical decision task; LNS = letter number sequencing test; NAL-R = National Acoustics Laboratory-revised version; MF = monaural far; MME = modified direct magnitude estimation; MN monaural near; N = number of participants; NC = near coding; NH = normal hearing; NoP = no processing; NR = noise reduction; NRDM = noise reduction and directional microphone; NV-hi = vocoded, high intelligibility; NV-lo = vocoded, low intelligibility; OHI = old, hearing impaired; ONH = old, normal hearing; OSLA matrix = Oldenburg sentences matrix test; P = primary task; pDTC = proportional -task; R = reverberation; Rspan = reading span; RSPIN = revised speech in noise test; RT = response time; S = secondary task; SIN = sentences in noise; SNR = signal to noise ratio; SPIN = speech in noise; SSN = speech spectrum noise; TLX = text load index; TVM sentences = closed-set sentence recognition test; TTB = 2-talker babble; VAP = visual-with concurrent auditory processing; VCV = vowel consonant vowel; VRT = visual response time; WAIS = Wechsler Adult Intelligence Scale; WM = working memory; WMC = working memory capacity; WMS = working memory span; YNH = young, normal hearing.

aUnilateral hearing loss was simulated by the use of an earplug with inserted into one ear during testing, resulting in a gradually sloping 4 dB per octave conductive hearing loss of approximately 30 dB HL at 500, 1,000, and 2,000 Hz.

bin basic mode, the hearing aid microphone was omnidirectional, and all advanced features, except for feedback management, were disabled, including directional processing and DNR. In advanced mode, the manufacturer's default settings for different listening environments were used. This included multichannel automatic directivity and algorithms designed to reduce reverberation, general background noise, and wind noise. The background noise level in this study (55 dBA) was not high enough to activate the devices' directional processing or noise-reduction algorithms. Thus, regardless of the aid setting (basic or advanced), the hearing aids were functioning in omnidirectional mode with no DNR active during testing. The interest was whether continuous access to advanced signal processing during daily activities would reduce cognitive processing demands and listening effort such that differences would be apparent when tested on the study's cognitively demanding dual task completed at the end of the day.

cNote on group labels: H+C+ = mild hearing loss, better cognitive function; H-C+ = moderate hearing loss, better cognitive function; H+C- = mild hearing loss, poorer cognitive function; H-C- = moderate hearing loss, poorer cognitive function

dNote that the study by Wild et al. describes a dual task and provides behavioral results for it, but the study did not examine listening effort or dual-task cost. Performance was only assessed for attended stimuli.

the audiovisual condition than in the auditory-alone condition. In contrast, Picou and coworkers (Picou et al., 2011, 2013) found that the provision of visual cues did not significantly increase listening effort for speech stimuli. Moreover, Picou et al. (2013) observed that better lip readers were more likely to derive benefit from the provision of visual cues and show a reduction in listening effort relative to an auditory-only condition. Clearly, more research is needed to account for the different findings reported by these two groups of investigators. One possible explanation may be the dual-task paradigm used; one group of investigators used a concurrent dual-task paradigm (Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010), whereas the other group used a sequential paradigm (Picou et al., 2011, 2013).

Relationship between dual-task outcomes and cognitive abilities. The relationship between the amount of listening effort expended during speech understanding and measures of cognitive abilities was assessed in several studies. For example, in some of the studies reviewed, the participants also performed a test of working memory capacity (WMC; Choi et al., 2008; Desjardins & Doherty, 2013; Neher et al., 2014; Neher, Grimm, Hohmann, & Kollmeier, 2014; Picou, 2011, 2013; Tun et al., 2009, 1991).

Desjardins and Doherty (2013) reported a significant correlation between outcomes on the Reading-Span test and listening effort in a group of participants who had hearing loss and who were fitted with hearing aids. Similarly, Tun et al. (1991) reported that WMC was a good predictor of listening effort. Picou et al. (2011) investigated listening effort for a speech recognition task administered in two modalities: auditory alone and audiovisual. The authors reported that participants who were good speech readers and who had a large WMC deployed less effort to perform the speech-understanding tasks. However, in other studies, the investigators failed to demonstrate that WMC was correlated to the listening effort displayed by adults with hearing loss who used hearing aids (Desjardins, 2016; Desjardins & Doherty, 2014; Neher et al., 2014; Neher, Grimm, Hohmann, et al., 2014). Furthermore, no associations with listening effort in listeners with a hearing loss were observed for a Stroop test of selective attention (Desjardins & Doherty, 2013) or the Digit Symbol Substitution Test for processing speed (Desjardins, 2016; Desjardins & Doherty, 2014). In summary, current results are inconclusive concerning the relationship between dual-task cost during speech understanding and WMC or other cognitive abilities.

Relationship between dual-task outcomes and self-report measures of listening effort. Questionnaires incorporating

Likert-like rating scales have been used to assess self-report listening effort. Two examples of such questionnaires are the *Quality of hearing* subscale of the “Speech, Spatial, and Qualities of Hearing” Scale (Gatehouse & Noble, 2004) and the Device-Oriented Subjective Outcome (Cox, Alexander, & Xu, 2014). Also, some investigators have used a unidimensional rating scale (e.g., ranging from *no effort* to *an extremely high level of effort*) to quantify the perceived listening effort during a listening task (Anderson-Gosselin & Gagné, 2011a, 2011b; Desjardins, 2016; Desjardins & Doherty, 2013, 2014; Fraser et al., 2010; Neher, Grimm, Hohmann, et al., 2014; Picou et al., 2011).

Several investigators have examined the relationship between behavioral and self-report measures of listening effort (or *ease of listening*; which can be considered the opposite of listening effort). In general, no associations between the two types of measures were observed (Anderson-Gosselin & Gagné, 2011a, 2011b; Desjardins, 2016; Desjardins & Doherty, 2013, 2014; Feuerstein, 1992; Fraser et al., 2010; Hornsby, 2013; Neher et al., 2014; Pals, Sarampalis, & Baskent, 2013; Picou et al., 2013). This consistent finding suggests that dual-task experimental paradigms may not measure the same attributes of listening effort as the concepts used by listeners when asked to rate the effort that was required to perform a specified speech-understanding task. Accordingly, Lemke and Besser (2016) have suggested that the term listening effort should be used as an umbrella term for *processing effort* (in terms of resource allocation) during listening on the one hand and for *perceived effort* (self-reported experience) on the other hand. McGarrigle et al. (2014) suggested that the term *dual-task cost* be used to describe the results obtained when a dual-task paradigm is used as a measure of listening effort.

Relationship between dual-task outcomes and other independent variables. This section summarizes how dual-task outcomes for listening effort are influenced by the characteristics of the masking signal, the SNR at which the speech task is administered, linguistic characteristics of the speech signal, and amplification or specific signal-processing algorithms incorporated into hearing aids.

Characteristics of the masker. In most of the studies reviewed, the speech material was presented in the presence of an interfering signal. In some of the studies, the interfering signal consisted of a masking noise (Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010). In two other studies, cafeteria noise was used as an interfering signal (Neher et al., 2014; Neher, Grimm, Hohmann, et al., 2014). Other investigators used speech as an interfering signal (e.g., Baer, Moore, & Gatehouse, 1993; Ng, Rudner, Lunner, & Rönnberg, 2015; Picou et al., 2011, 2013; Xia et al., 2015). In one

of the studies, the interfering signal consisted of backward speech (Helfer et al., 2010). A thorough discussion on the advantages and disadvantages of using nonspeech signals as an interfering signal (i.e., purely energetic masking) versus speech stimuli (which includes both energetic and semantic or informational masking) is beyond the scope of the present review. Several investigators have discussed these two types of masking signals in relation to performance on speech-understanding tasks (Brungart, Simpson, Ericson, & Scott, 2001; Freyman, Balakrishnan, & Helfer, 2004; Kidd, Mason, Richards, Gallun, & Durlach, 2008; Rosen, Souza, Ekelund, & Majeed, 2013).

Desjardins and Doherty (2013) found that the type of the masking signal had an influence on listening effort but the effect was different for younger and older listeners. Helfer et al. (2010) found that listening effort is influenced by the spatial configuration of the masking source relative to the target speech. Xia et al. (2015) reported that the availability of spatial separation cues or voice-difference cues also has an effect on listening effort.

SNR at which the speech task is administered. In some studies, the effect of the SNR on dual-task outcomes was tested (Baer et al., 1993; Neher et al., 2014; Neher, Grimm, Hohmann, et al., 2014; Sarampalis, Kalluri, Edwards, & Hafter, 2009). In all of these studies, it was found that performance on the secondary task improved (i.e., the dual-task cost decreased) when the SNR was improved. However, it is noteworthy that in those studies, performance on the primary task also improved as the SNR improved. Hence, it is not possible to rule out the possibility that the reduction in dual-task cost on the secondary task was attributable to the fact that the primary task became less difficult as the SNR was improved.

A related issue is the SNR at which the speech-understanding task is administered. In most of the studies reviewed, the SNR selected was such that it yielded a high-level speech-understanding accuracy score (e.g., 80% correct), while avoiding ceiling effects. In a recent study Wu et al. (2016) used a dual-task paradigm to investigate listening effort as a function of the SNR at which the primary speech recognition task was administered. For each experimental condition of the primary task, RTs to both an easy task (i.e., visual probe detection) and a difficult task (i.e., a color Stroop test) were used to measure performance for the secondary task. Three different (but related) experiments yielded a similar pattern of results concerning the secondary task. Specifically, RTs for the secondary task had a maximum duration when the SNR was set so that the primary-task performance level yielded speech recognition scores that were in the range of 30% to 50% correct (i.e., at SNRs of

–2 and 0 dB). RTs for the secondary task were shorter when the SNR was set to yield either lower or higher speech recognition scores. The unexpected finding in these experiments is that the maximum duration RT was not observed under the experimental condition in which the sentences were presented at the poorest SNRs. The findings reported by Wu et al. (2016) suggest that the experimental condition under which listening effort is most likely to be revealed occurs when the primary speech-perception task is designed to results in a performance level that gives a percent correct response rate between 30% and 50%.

Linguistic characteristics of the speech signal. In some studies, the revised Speech in Noise Test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Wilson, McArdle, Watts, & Smith, 2012) was used to assess listening effort using a dual-task paradigm (Desjardins & Doherty, 2013, 2014; Feuerstein, 1992; Sarampalis et al., 2009). The results of these investigations showed that listening effort was greater for the low-predictability sentences than for the high-predictability sentences.

Amplification or specific types of signal processing. A number of investigators have reported the effects of amplification or specific types of signal-processing algorithms on listening effort. Overall, regardless of the type of dual-task paradigm used, listening effort appears to decrease with the use of (a) amplification (Baer et al., 1993; Downs, 1982; Hornsby, 2013; Picou et al., 2013), (b) dynamic-range compression (Baer et al., 1993), (c) noise-reduction algorithms (Desjardins & Doherty, 2014; Sarampalis et al., 2009), (d) directional microphones (Wu et al., 2014), (e) advanced compared with basic hearing-aid settings (Hornsby, 2013), and (f) combinations of directional microphones and noise reduction (Desjardins, 2016). However, it should be noted that sometimes, changes in listening effort were observed only for some of the experimental conditions tested. In fact, in two studies, it was found that the use of an aggressive noise-reduction algorithm resulted in an increase in listening effort (Neher et al., 2014; Neher, Grimm, Hohmann, et al., 2014). At the present time, it is not clear whether the differences in findings across the studies are due to differences in the noise-reduction algorithms used, differences in the dual-task paradigms employed, differences in other components of the experimental procedures, or differences in the characteristics of the participants.

Methodological Issues and Design Considerations

While analyzing the studies retained for the present review, it became evident that several decisions made during the design stage of investigations incorporating

a dual-task paradigm may influence the results obtained. This section reviews some of the identified methodological issues.

Type of dual-task paradigm (concurrent vs. sequential). Under the dual-task condition, within a single test trial, the primary task and the secondary task may be administered concurrently or sequentially. With concurrent stimulus presentation, the participant is required to process the stimuli of both tasks at the same time, such as in the studies by Desjardins and colleagues (Desjardins, 2016; Desjardins & Doherty, 2013, 2014). The secondary task in those studies, initially described by Kemper, Schmalzried, Herman, Leedahl, and Mohankumar (2009), consisted of the digital pursuit rotor task, a visual motor tracking task performed with a computer mouse. Under the dual-task condition, the participants had to listen to and repeat the primary-task sentences they heard while performing the digital pursuit rotor task.

In sequential dual-task paradigms, a period of time separates the presentation of the primary- and the secondary-task stimuli, such as described by Rakerd et al. (1996). In that study, a list of digits appeared on a computer monitor. The participants were instructed to memorize the list for later recall. Then, the participants heard a short text of connected discourse that was approximately 1 minute in duration and answered questions about the information presented. Afterwards, the participants were asked to recall the list of digits initially presented to them. More recently, a similar sequential dual-task paradigm was employed in some studies reported by Picou et al. (2011, 2013).

Among the studies conducted with adult participants, an overwhelming majority of the studies used a concurrent experimental paradigm. In fact, only one of the studies reviewed included a sequential design (Rakerd et al., 1996). It would appear as though there is a strong bias to use concurrent dual-task paradigms to investigate listening effort for speech. One of the reasons for this tendency may be that investigators wish to capitalize on the fact that requesting a participant to perform two tasks concurrently holds a high level of ecological validity because multitasking is something that persons are often required to do as part of their everyday life activities. Furthermore, concurrent dual tasking taps into more processing resources than merely memory functions.

Primary tasks of dual-task paradigms. In most of the studies reviewed, the primary task was sentence recognition (Anderson-Gosselin & Gagné, 2011a, 2011b; Baer et al., 1993; Desjardins, 2016; Desjardins & Doherty, 2013, 2014; Feuerstein, 1992; Fraser et al., 2010; Helfer et al., 2010; Neher, Grimm, Hohmann, et al., 2014; Pals et al., 2013; Sarampalis et al., 2009; Wild et al., 2012; Wu

et al., 2014). However, in one study, the primary task was syllable recognition (Rigo, 1986), and in other studies, the primary task was recognition of words presented in isolation (Downs, 1982; Hornsby, 2013; Picou & Ricketts, 2014; Picou et al., 2011, 2013; Tun et al., 2009). In two studies, the stimuli used were spoken passages, and the task consisted of a speech-comprehension response (Rakerd et al., 1996; Tun et al., 1991). Some studies used a closed-set response task, whereas other investigators used an open-set speech recognition task. One advantage of using a closed-set recognition task is that it makes it easy to determine when the participant initiates response (i.e., key-stroke or finger response on a touch screen monitor) and thus to assess the RT as well as accuracy.

Secondary tasks used. The secondary tasks that have been used in dual-task studies of listening effort for speech understanding are quite diverse (see Table 2). In most instances, the investigators did not motivate their choice of secondary task. Secondary tasks can be accuracy or RT measures. Seemingly, it is assumed that the type of secondary task and the type of evaluation metric used will not influence the measurement of listening effort.

Presently, no specific category of secondary task has been shown to be the most appropriate for measuring listening effort using a dual-task paradigm. Picou and Ricketts (2014) reported two experiments in which the goal of the study was to evaluate the effects of the secondary task on listening-effort outcomes. In each experiment, three different secondary tasks, based on RT measures, were compared: (a) simple visual probe, (b) complex visual probe, and (c) category of word presented. The results of Experiment 1 (conducted with participants with normal hearing) and Experiment 2 (conducted with participants with a mild-to-moderate hearing loss) revealed that only the secondary task that required a word-category recognition response was sensitive to changes in the primary task. According to the investigators, word-category recognition required *deeper* processing than the two visual-probe tasks. It is difficult to determine whether the secondary task that consisted of the word-category recognition task was more sensitive to listening effort because it involved processing linguistic information (just as the primary task did) or because it was cognitively more demanding in other respects. Nevertheless, the results of the investigation do suggest that the type of secondary task used may have an influence on whether or not listening effort can be measured.

In another study (Wu et al., 2014), different secondary tasks were used in two companion experiments in which the same participants were involved. In one experiment, the secondary task consisted of driving a vehicle in a driving simulator, and in the other study, it consisted

Table 2. List of Types of Secondary Tasks That Were Used in the Reviewed Dual-Task Studies on Listening Effort for Speech.

Visual motor tracking task displayed on a computer screen	(Desjardins, 2016; Desjardins & Doherty, 2013, 2014; Tun et al., 2009; Xia et al., 2015)
Response time to visual probe or distractor	(Downs, 1982; Feuerstein, 1992; Hornsby, 2013; Neher et al., 2014; Neher, Grimm, Hohmann, et al., 2014; Picou & Ricketts, 2014; Picou et al., 2013; Sarampalis et al., 2009; Tun et al., 1991; Wild et al., 2012; Wu et al., 2014)
Recall of words presented in the primary task or before the primary task	(Hornsby, 2013; Ng et al., 2015; Picou et al., 2011; Rakerd et al., 1996; Sarampalis et al., 2009)
Judgment concerning a feature of the primary task	(Baer et al., 1993; Helfer et al., 2010)
Tactile pattern recognition	(Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010)
Semantic judgment task	(Picou & Ricketts, 2014)
Driving a car simulator	(Wu et al., 2014)
Mental rotation of Japanese characters	(Pals et al., 2013)
Response time to predetermined auditory signals	(Wild et al., 2012)
Rhyme judgment for words presented visually	(Pals et al., 2013, 2015)

of a visual-pattern recognition task. In both experiments, performance on the secondary task revealed a significant effect of listening effort. Moreover, the performance levels on both tasks were found to be significantly correlated with each other. This finding supports the idea that different categories of secondary tasks may be equally appropriate to measure listening effort using a dual-task paradigm. Notwithstanding these observations, at the present time it would be unwise to conclude that *any* secondary task can be used to investigate listening effort for speech understanding and that all secondary tasks are equally sensitive to differences in listening effort. In summary, further research may be required to identify which type of secondary task or which combination of primary and secondary tasks are best suited for investigating listening effort for speech understanding.

Ways of quantifying listening effort

Proportion correct responses versus RT measures of listening effort. In some studies, the dependent variable used to quantify listening effort consisted of the proportion of correct scores, while in other studies, it consisted of a RT measure (refer to Table 1 for an overview of dependent measures used across studies). At the present time, there is no clear indication concerning which of the two approaches is the most appropriate to characterize listening effort. However, it is noteworthy that Whelan (2008) claims that in some experiments, the RT data collected may not meet some of the assumptions required to use an analysis of variance to test for significant effects.

Some investigators have chosen to compute both proportion correct scores and RTs for the secondary task

(Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010). For example, Gagné and his colleagues (Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010) used a two-element tactile pattern recognition task as their secondary task. The four possible response alternatives were displayed on a touch screen computer monitor. The participants were instructed to respond as quickly as possible by selecting their choice of response from the alternatives shown on the monitor. Both accuracy and RTs were assessed. As a first approximation, both methods used to quantify performance on the secondary task led to a similar pattern of results.

The metric used to measure listening effort. The classic method used to quantify listening effort consists of calculating the difference score (DS) between the baseline and the dual-task performance on the secondary task (i.e., $\text{Listening effort} = \text{Secondary task}_{\text{baseline}} - \text{Secondary task}_{\text{dual-task}}$). However, a simple DS should not be used if there is a large difference between groups in the baseline secondary-task performance. For example, a DS of 30 ms should be interpreted differently, when the baseline RT was 540 ms (i.e., $540 \text{ ms} - 510 \text{ ms} = 30 \text{ ms}$) compared with when the baseline RT was 80 ms (i.e., $80 \text{ ms} - 50 \text{ ms} = 30 \text{ ms DS}$). One way to circumvent this issue is to use a proportional DS. For example, Fraser et al. (2010) computed a proportional dual-task cost (pDTC) as their measure of listening effort, where $\text{pDTC} = \frac{\text{Secondary-task}_{\text{baseline}} - \text{Secondary-task}_{\text{dual-task}}}{\text{Secondary-task}_{\text{baseline}}} \times 100$.

Similarly, there may be a difference in a participant's performance level on the primary task when it is

administered under the baseline condition and when it is administered concurrently under the dual-task condition. In such cases, it may be appropriate to compute a pDTC for the primary task in addition to a pDTC for the secondary task. Significant differences in pDTC for either the primary task or the secondary task could be interpreted as an indication of a significant effect of listening effort (e.g., Fraser et al., 2010). Moreover, when pDTCs are computed for both the primary and the secondary tasks, it is justified to compute an aggregate listening effort index by combining the listening effort scores (e.g., pDTC) obtained for both the primary and the secondary tasks. For example, one may choose to report the pDTC_{total} score (whereby, the $\text{pDTC}_{\text{total}} = \text{pDTC}_{\text{primary task}} + \text{pDTC}_{\text{secondary task}}$). Notably, an aggregate listening effort index was not reported in any of the studies included in the present review. However, a rationale in support of using an approach similar to the one described here was recently provided by Plummer and Eskes (2015). The investigators claim that the magnitude and direction of dual-task interference may be influenced by the interaction between the two tasks and by how individuals spontaneously prioritize their attention. The authors demonstrate that this approach to measuring dual-task cost takes into account the trade-offs in performance that the participant may attribute to the primary and the secondary tasks.

Same experimental condition versus same level of performance. When a study is designed to compare listening effort across different groups of participants, the investigator must decide whether the speech-understanding task should be administered under the same conditions across groups (e.g., at the same SNR) or at equal speech-understanding baseline performance (e.g., using a different SNR for each group). Both options may be appropriate, depending on the research question that is being addressed. The same issue also applies when different experimental conditions (e.g., Amplification System A vs. Amplification System B) are administered to the same group of participants, that is, when the study consists of a within-subject design. One alternative could be to incorporate both experimental setups in the same study, as was done in some of the reviewed studies (e.g., Anderson-Gosselin & Gagné, 2011a, 2011b; Fraser et al., 2010). This latter alternative makes it possible to address the results from both an ecological (everyday life situation) as well as a conceptual (measuring listening effort at the same level of performance) standpoint.

Other methodological considerations. For a secure interpretation of differences in secondary-task outcomes in terms of listening effort, it is required that the listener is motivated to aim at high levels of performance in both primary task and secondary task, as well as that the tasks

are challenging enough to draw on the full investment of required processing resources available in the listener's processing system. Otherwise, when no difference in secondary-task outcome is observed, one does not know whether there was no difference in effort for the different primary-task conditions or whether there were sufficient resources left for performing a more effortful primary task without compromising performance on the secondary task. This methodological requirement is hard to ensure, given that we have no means to assess the general individual capacity available.

Another methodological concern is that priority should be given to the primary task by the listener under all circumstances for an easy interpretation of the experimental results. However, it may not always be possible to monitor whether this condition is satisfied. For example, a dual-task study conducted with children revealed that instructions given to participants on how to prioritize the two tasks were ineffective (Choi et al., 2008).

Discussion and Conclusions

The current article presents an overview of previous studies that have used dual-task paradigms to assess listening effort during speech understanding. We would like to stress that while we strove to perform a scoping review, the review was not performed according to guidelines for a systematic review and did not aim to evaluate the quality of the included studies in any way. Rather, the purpose was to thoroughly describe the previous publications to provide an overview of what has been done so far, given that dual tasks for measuring listening effort are a relatively new field within hearing research. Specifically, the aims of this review were to (a) describe the large variety of methodological approaches that have been applied, especially the plenty of secondary tasks; (b) provide a broad summary of the results that have been obtained, especially regarding effects of listener age, hearing status, and signal processing; (c) discuss a number of methodological considerations that need to be taken into account when designing a dual-task study.

The present review revealed a large variability in the experimental paradigms used to behaviorally measure listening effort during speech understanding, in terms of the primary and secondary tasks applied as well as the experimental manipulations for which effects on effort were assessed, and the listener groups. While most of the applied paradigms were able to detect changes in listening effort related to an experimental manipulation, this large variability in applied settings makes it difficult to draw any firm conclusions about the most suitable dual-task paradigm for assessing listening effort or about the sensitivity of specific paradigms to different types of experimental manipulations. Overall,

systematic evaluations of the applied paradigms, including psychometric properties, are lacking and would be highly desired.

Specifically, the relationship between dual-task cost and the following factors should be examined. First, in the vast majority of the studies, the SNR for the primary speech-understanding task was set in such a way that the participants obtained a high level of baseline accuracy performance, often in the range of 80% correct or better. It would be of interest to examine the relationship between the level of performance on the primary-task baseline measure and the dual-task cost. For example, is there a linear relationship between primary-task baseline and dual-task cost on the secondary task, such that doubling accuracy on the primary task leads to halving dual-task costs on the secondary task?

Second, it is not known whether the additivity of different sources of listening effort is linear or nonlinear. For example, the dual-task cost of performing a speech recognition task in noise for individuals with a moderately severe hearing loss may be equal to the dual-task cost of performing the same speech task for persons with normal hearing who perform the task in a nonnative language. What is the dual-task cost expected for individuals with a moderately severe hearing loss who perform the speech recognition task in their nonnative language?

Third, whereas some investigators (Picou & Ricketts, 2014; Wu et al., 2014) have compared the effects of using different secondary tasks when assessing listening effort, to date there is no general understanding of the suitability and sensitivity of the different types of secondary tasks. Is a secondary task that calls upon the use of short-term memory as sensitive as a secondary task that calls upon the use of visual attention skills? Recently, Kahneman's (1973) capacity model of attention has been adapted to the specific case of listening effort, resulting in the framework for understanding effortful listening (FUEL; Pichora-Fuller et al., 2016). The FUEL framework consolidates the general assumptions of the Kahneman's model that resources are limited and shared between tasks and illustrates how effort is influenced by interactions of external task demands and internal motivation. Accordingly, the FUEL framework confirms the theoretical assumptions underlying the dual-task approach. However, to our knowledge, neither the FUEL framework nor any other model of cognitive resources (cf. Wingfield, 2016) provides information that would theoretically motivate the choice of a specific type of secondary task. For example, it may seem obvious that a secondary task in the same modality (auditory) as the primary task or the same processing domain (verbal) would compete more with speech understanding than other tasks. Nonetheless, as described in the present review, dual-task costs during

speech understanding have also been found for secondary tasks such as motor tracking and tactile pattern recognition. It has been described that multitasking leads to an increased activation of and demand for executive functions (Diamond, 2013). Possibly, the reallocation of resources toward executive functions in dual tasking is more important than the domain or modality of the secondary task and potentially also than the specific design of the secondary task.

Fourth, both sequential and concurrent dual-task paradigms can be used to measure listening effort. At the present time, it is not known if these two approaches measure the same dimensions of listening effort. If not, then which approach is best suited to measure listening effort for speech-understanding tasks?

Finally, the participant is generally instructed to optimize performance on the primary task rather than on the secondary task. Data obtained from children suggest that this simple instruction may not be sufficient to ensure that the participant will optimize performance on the primary task (Choi et al., 2008; Irwin-Chase & Burns, 2000). It would be of interest to examine the same issue in adults. Also, in a previous section, it was suggested that perhaps one way of overcoming this issue may be to compute an aggregate dual-task cost. However, at the present time, there are no empirical data available to support the use of this strategy.

In sum, the present review provides a comprehensive and critical analysis of investigations in which a dual-task paradigm was used to investigate aspects of speech understanding among younger and older adults. Generally, the results of the analysis suggest that this type of experimental procedure appears to be sensitive to a number of differences in experimental conditions, both across groups of participants as well as within the same group of listeners. At the same time, systematic evaluations of the existing paradigms are needed for making informed design decisions. Given the importance of attentional and other cognitive processes involved in speech understanding, it would be of interest to pursue investigations that contribute to the development of a clinical procedure that will make it possible to quantify the listening effort during speech understanding. In the long term, the use of dual-task experimental paradigms may constitute a good approach toward achieving this goal.

The review also revealed that at the present time, there does not appear to be a consensus on the type of dual-task experimental paradigm that is most appropriate to investigate listening effort for speech understanding. Several differences in experimental procedures were apparent across investigations. Also, several issues that warrant further investigations were identified. The present review should be of interest to investigators who are interested in applying dual-task experimental paradigms

to investigate issues related to listening effort. Finally, several research questions that warrant further investigation in order to better understand and characterize the intricacies of dual-task paradigms were proposed.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Anderson-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. doi:10.3109/14992027.2011.599870.
- Anderson-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. doi:10.1044/1092-4388(2010/10-0069).
- Baer, T., Moore, B. C., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development*, *30*, 49–49.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeckowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, *27*(1), 32–48. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6717005>.
- Broadbent, D. (1958). *Perception and communication*. London, England: Pergamon Press.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5 Pt 1), 2527–2538. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11757942>.
- Choi, S., Lotto, A., Lewis, D., Hoover, B., & Stelmachowicz, P. (2008). Attentional modulation of word recognition by children in a dual-task paradigm. *Journal of Speech, Language, and Hearing Research: JSLHR*, *51*(4), 1042–1054. doi:10.1044/1092-4388(2008/076).
- Cox, R. M., Alexander, G. C., & Xu, J. (2014). Development of the device-oriented subjective outcome (DOSO) scale. *Journal of the American Academy of Audiology*, *25*(8), 727–736. doi:10.3766/jaaa.25.8.3.
- Desjardins, J. L. (2016). The effects of hearing aid directional microphone and noise reduction processing on listening effort in older adults with hearing loss. *Journal of the American Academy of Audiology*, *27*(1), 29–41. doi:10.3766/jaaa.15030.
- 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. doi: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. doi:10.1097/AUD.0000000000000028.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. doi:10.1146/annurev-psych-113011-143750.
- Downs, D. W. (1982). Effects of hearing and use on speech discrimination and listening effort. *The Journal of Speech and Hearing Disorders*, *47*(2), 189–193. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7176597>.
- Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear and Hearing*, *13*(2), 80–86.
- 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: JSLHR*, *53*(1), 18–33. doi:10.1044/1092-4388(2009/08-0140).
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5 Pt 1), 2246–2256. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15139635>.
- Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*, 85–99. doi:10.1080/14992020400050014.
- Helfer, K. S., Chevalier, J., & Freyman, R. L. (2010). Aging, spatial cues, and single- versus dual-task performance in competing speech perception. *The Journal of the Acoustical Society of America*, *128*(6), 3625–3633. doi:10.1121/1.3502462.
- 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.
- Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523–534. doi:10.1097/AUD.0b013e31828003d8.
- Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology*, *49*(12), 928–932.
- Hughes, K. C., & Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear implants international*, *14*(3), 121–129.
- Irwin-Chase, H., & Burns, B. (2000). Developmental changes in children's abilities to share and allocate attention in a dual task. *Journal of Experimental Child Psychology*, *77*(1), 61–85.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

- Kemper, S., Schmalzried, R., Herman, R., Leedahl, S., & Mohankumar, D. (2009). The effects of aging and dual task demands on language production. *Aging, Neuropsychology, and Cognition, 16*(3), 241–259.
- Kidd, G. Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–189). Berlin, Germany: Springer.
- Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and Hearing, 37*(Suppl 1): 77S–84S.
- 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), 433–440. doi:10.3109/14992027.2014.890296.
- McFadden, B., & Pittman, A. (2008). Effect of minimal hearing loss on children’s ability to multitask in quiet and in noise. *Language, speech, and hearing services in schools, 39*(3), 342–351.
- Neher, T., Grimm, G., & Hohmann, V. (2014). Perceptual consequences of different signal changes due to binaural noise reduction: Do hearing loss and working memory capacity play a role? *Ear and Hearing, 35*(5), e213–e227. doi:10.1097/AUD.0000000000000054.
- Neher, T., Grimm, G., Hohmann, V., & Kollmeier, B. (2014). Do hearing loss and cognitive function modulate benefit from different binaural noise-reduction settings? *Ear and Hearing, 35*(3), e52–e62. doi:10.1097/AUD.0000000000000003.
- Ng, E. H., Rudner, M., Lunner, T., & Rönnerberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear and Hearing, 36*(1), 82–91. doi:10.1097/AUD.0000000000000080.
- Pals, C., Sarampalis, A., & Baskent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research: JSLHR, 56*(4), 1075–1084. doi:10.1044/1092-4388(2012/12-0074).
- Pals, C., Sarampalis, A., van Rijn, H., & Baskent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America, 138*(3), EL187. doi:10.1121/1.4929614.
- Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society, 23*(10), 433–441.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... Wingfield, A. D. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*(Suppl 1): 5S–27S.
- Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and hearing, 37*(1), 1–13.
- 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. doi:10.1097/AUD.0000000000000055.
- Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research: JSLHR, 54*(5), 1416–1430. doi:10.1044/1092-4388(2011/10-0154).
- Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing, 34*(5), e52–e64. doi:10.1097/AUD.0b013e31827f0431.
- Plummer, P., & Eskes, G. (2015). Measuring treatment effects on dual-task performance: A framework for research and clinical practice. *Frontiers in Human Neuroscience, 9*, 225.
- Rakerd, B., Seitz, P. F., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing, 17*(2), 97–106. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8698163>.
- Rigo, T. G. (1986). The relationship between the visual contribution to speech perception and lip reading ability during focused and divided attention. *Ear and Hearing, 7*(4), 266–272.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Pichora-Fuller, M. K. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*, 31.
- 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*(S2), S99–S105.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America, 133*(4), 2431–2443. doi:10.1121/1.4794379.
- 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.
- 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.
- Stelmachowicz, P. G., Lewis, D. E., Choi, S., & Hoover, B. (2007). The effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children. *Ear and hearing, 28*(4), 483.
- Tun, P. A., McCoy, S. L., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging, 24*(3), 761.
- Tun, P. A., Wingfield, A., & Stine, E. A. (1991). Speech-processing capacity in young and older adults: A dual-task study. *Psychology and Aging, 6*(1), 3–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2029365>.
- Wechsler, D. (1997). *Wechsler memory scale (WMS-III)*. San Antonio, TX: Psychological Corporation.

- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475.
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *The Journal of Neuroscience*, 32(40), 14010–14021.
- Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The Revised Speech Perception in Noise Test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *Journal of the American Academy of Audiology*, 23(8), 590–605. doi:10.3766/jaaa.23.7.9.
- Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing*, 37(Suppl 1): 35S–43S. doi:10.1097/AUD.0000000000000310.
- Wu, Y. H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., & Bentler, R. (2014). Measuring listening effort: Driving simulator versus simple dual-task paradigm. *Ear and Hearing*, 35(6), 623–632. doi:10.1097/AUD.0000000000000079.
- Wu, Y. H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear and Hearing*, 37(6), 660–670.
- Xia, J., Nooraei, N., Kalluri, S., & Edwards, B. (2015). Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 137(4), 1888–1898. doi:10.1121/1.4916599.