



Published in final edited form as:

Ear Hear. 2016 ; 37(6): 660–670. doi:10.1097/AUD.0000000000000335.

Psychometric functions of dual-task paradigms for measuring listening effort

Yu-Hsiang Wu¹, Elizabeth Stangl¹, Xuyang Zhang¹, Joanna Perkins¹, and Emily Eilers¹

¹Department of Communication Sciences and Disorders, The University of Iowa

Abstract

Objectives—The purpose of the study was to characterize the psychometric functions that describe task performance in dual-task listening effort measures as a function of signal-to-noise ratio (SNR).

Design—Younger adults with normal hearing (YNH, $n = 24$; Experiment 1) and older adults with hearing impairment (OHI, $n = 24$; Experiment 2) were recruited. Dual-task paradigms wherein the participants performed a primary speech recognition task simultaneously with a secondary task were conducted at a wide range of SNRs. Two different secondary tasks were used: an easy task (i.e., a simple visual reaction-time task) and a hard task (i.e., the incongruent Stroop test). The reaction time (RT) quantified the performance of the secondary task.

Results—For both participant groups and for both easy and hard secondary tasks, the curves that described the RT as a function of SNR were peak shaped. The RT increased as SNR changed from favorable to intermediate SNRs, and then decreased as SNRs moved from intermediate to unfavorable SNRs. The RT reached its peak (longest time) at the SNRs at which the participants could understand 30% to 50% of the speech. In Experiments 1 and 2 the dual-task trials that had the same SNR were conducted in one block. To determine if the peaked shape of the RT curves was specific to the blocked SNR presentation order used in these experiments, YNH participants were recruited ($n = 25$; Experiment 3) and dual-task measures, wherein the SNR was varied from trial to trial (i.e., non-blocked), were conducted. The results indicated that, similar to the first two experiments, the RT curves had a peaked shape.

Conclusions—Secondary task performance was poorer at the intermediate SNRs than at the favorable and unfavorable SNRs. This pattern was observed for both YNH and OHI participants and was not affected by either task type (easy or hard secondary task) or SNR presentation order (blocked or non-blocked). The shorter RT at the unfavorable SNRs (speech intelligibility < 30%) possibly reflects that the participants experienced cognitive overload and/or disengaged

Address correspondence to: Yu-Hsiang Wu, 125C SHC, Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, IA 52242, USA. yu-hsiang-wu@uiowa.edu, Fax number: 319-335-8851, Telephone number: 319-335-8728.

Portions of this paper were presented at the annual meeting of the American Auditory Society, March, 2013, Scottsdale, Arizona, USA, and the annual ASHA Convention, November, 2015, Denver, Colorado, USA.

Conflicts of Interest and Source of Funding:

Yu-Hsiang Wu is currently receiving grants from NIH and the National Institute on Disability and Rehabilitation Research. For the remaining authors, none were declared. Experiments 1 and 2 were supported by the New Century Scholars Research Grant from American Speech-Language-Hearing Foundation. Experiment 3 was supported by NIH/NIDCD R03DC012551 and R01-DC012769. The pilot study of this project was supported by a research grant from Siemens Hearing Instruments.

themselves from the listening task. The implication of using the dual-task paradigm as a listening effort measure is discussed.

Keywords

listening effort; dual-task paradigm

INTRODUCTION

Understanding speech involves both auditory and cognitive factors (e.g., Kiessling et al. 2003; Worrall & Hickson 2003; Pichora-Fuller & Singh 2006; Humes 2007). Listening effort, which is defined as “the mental exertion required to attend to, and understand, an auditory message” (McGarrigle et al. 2014), has been recognized as an important dimension of speech perception and hearing enhancement device outcomes. When the target auditory message is speech, listening effort is often conceptualized as the cognitive resources allocated for speech processing (Hick & Tharpe 2002; Fraser et al. 2010; Gosselin & Gagné 2010; Zekveld et al. 2010).

Among different methodologies that can objectively quantify listening effort, the dual-task paradigm is one of the most widely-used behavioral measures (Gosselin & Gagné 2010). In this paradigm, listeners perform a primary speech recognition task concurrently with a secondary task. The former is referred to as the primary task because listeners are instructed to maximize speech recognition performance. The difficulty of the speech recognition task is systematically varied during the test session and the change in secondary task performance is taken as an index of the shift in allocation of cognitive resources for speech processing, i.e., listening effort. This interpretation assumes that (1) performance on each of the tasks requires some common cognitive resource allocation and (2) cognitive resources are limited (Kahneman 1973). Dual-task paradigms have been used to investigate the effect of age (Gosselin & Gagné 2011; Desjardins & Doherty 2013; Degeest et al. 2015), hearing loss (Hick & Tharpe 2002), visual cues (Fraser et al. 2010; Picou et al. 2013; Picou & Ricketts 2014), hearing aids (Downs 1982; Hornsby 2013), noise reduction algorithms (Sarampalis et al. 2009; Desjardins & Doherty 2014), and directional microphones (Wu et al. 2014) on listening effort.

Although dual-task tests are widely used, it is less clear at what speech intelligibility level the test will be most sensitive to changes in listening effort (i.e., changes in secondary task performance). Specifically, researchers have speculated that dual-task measures conducted in conditions wherein the primary speech recognition task is too easy (e.g., quiet) or too difficult (e.g., high-level background noise) may not be sensitive (e.g., Picou et al. 2013). Gatehouse and Gordon (1990) measured listening effort using the reaction time (RT) to speech stimuli. They examined the effect of hearing aid amplification on listening effort in four conditions: when the unaided speech intelligibility was between 15 to 85%, close to 50%, and close to 85%, and when the benefit of amplification on speech intelligibility was less than 6%. These researchers found that their listening effort measure was more sensitive in the first two conditions. No prior study, however, has systematically examined the

relationship between speech intelligibility and secondary task performance in dual-task listening effort measures.

To fill this gap, the objective of the current research was to characterize the psychometric functions that describe task performance in dual-task listening effort measures as a function of signal-to-noise ratio (SNR). It was hypothesized that speech recognition performance would decrease monotonically as SNR decreases. Based on the Ease of Language Understanding (ELU) model that describes a conceptual framework for speech understanding (Rönnberg 2003; Rönnberg et al. 2008; Rönnberg et al. 2013), it was further hypothesized that secondary task performance would decrease monotonically as SNR (and speech intelligibility) decreases and might reach an asymptote level at very poor SNRs. Specifically, the ELU model suggests that speech information is rapidly and automatically bound in an episodic buffer and then compared to long-term memory. If the information in the buffer matches the long-term memory, speech will be recognized and there is no need for top-down processing. If the speech information input cannot immediately match the long-term memory, explicit and deliberate working memory top-down processes will be invoked to compensate for this mismatch. Because at poorer SNRs speech information will be more degraded and the mismatch will be more likely to happen, more top-down processing will be recruited for speech recognition and less processing will be available to the secondary task, resulting in poorer secondary task performance.

This hypothesis is supported by empirical data showing that the cognitive processing load of a task increases monotonically as the task becomes more demanding (Peavler 1974; Cabestrero et al. 2009; Zekveld & Kramer 2014). For example, Cabestrero et al (2009) measured the cognitive processing load of an auditory digit span recall task using pupillometry. The number of the to-be-recalled digits was manipulated to create three load conditions (5, 8, and 11 digits). The results showed that pupil response systematically increased as more digits were presented. In the most difficult 11-digit condition, the pupil response reached an asymptote level at the ninth digit and remained stable until the last digit was presented. More recently, Zekveld and Kramer (2014) used pupil response to measure the cognitive processing load of a speech recognition task from a group of younger adults with normal hearing. The experiment consisted of one quiet and three noisy conditions (the first experiment of Zekveld & Kramer 2014). The mean speech intelligibility of the four conditions was 99%, 94%, 54%, and 0%, respectively. The result showed that pupil response increased linearly as SNR and speech intelligibility decreased.

There is evidence, however, suggesting that secondary task performance of dual-task listening effort measures may not decrease monotonically as SNR decreases (Poock 1973; Granholm et al. 1996; Zekveld & Kramer 2014). Granholm et al (1996) measured pupil response of an auditory digit span recall task in three load conditions (5, 9, and 13 digits). The 5- and 9-digit conditions were low- and moderate-load conditions, respectively, and the 13-digit condition was considered overload as it exceeded the memory span and was almost impossible to recall completely. The study results showed that pupil response, instead of increasing monotonically as the recall task became more demanding, was smaller in the 13-digit condition than in the 5- and 9-digit conditions. Similar findings were observed by a second Zekveld and Kramer (2014) experiment where a speech recognition test was

administered at a wide range of SNRs. The experiment results indicated that as SNR decreased from -4 dB (speech intelligibility: 80%) to -8 dB (intelligibility: 50%), pupil response increased. As SNR further decreased from -8 dB to -36 dB (intelligibility: 0%), however, pupil response decreased linearly. In other words, the function that described pupil response across SNRs had a peaked shape. Zekveld and Kramer (2014) suggested that their participants experienced cognitive overload (i.e., task demands exceed an individual's ability) at poorer SNRs, as did the subjects in the overload condition of the study by Granholm et al (1996). It is possible that people in the overload condition tend to give up on the task. Zekveld and Kramer (2014) asked subjects to report how often they gave up listening to speech and found that the frequency of giving up increased as speech intelligibility decreased.

The research finding showing that pupil response is smaller in the very demanding, overload condition than in the moderate-load condition is consistent with a study by Petersen et al (2015), who measured working memory load using alpha oscillations of electroencephalogram. In the experiment, older adults with various degrees of hearing loss wore hearing aids and were asked to conduct an auditory recall task. The results showed that in low- and moderate-load conditions alpha power increased as the degree of hearing loss increased. In the most demanding condition, however, alpha power of subjects with moderate hearing loss was lower than that of subjects with mild hearing loss. Similarly, Sander et al. (2012) found that alpha power in high demanding conditions is reduced in older adults compared to children and younger adults. Therefore, this line of research suggests that (1) the psychometric function of secondary task performance in the dual-task listening effort measure would have a peaked shape (i.e., performance being poorer at intermediate SNRs than at favorable and unfavorable SNRs) and (2) the functions of younger adults with normal hearing (YNH) and older adults with hearing impairment (OHI) would have different shapes.

The current research consisted of three experiments. The first two experiments characterized the psychometric functions of the dual-task paradigm for YNH (Experiment 1) and OHI (Experiment 2) listeners. To determine if the trend of task performance across SNRs found in Experiments 1 and 2 was specific to the methodology used in the dual-task measures, Experiment 3 was conducted on YNH listeners using a different methodology. For all three experiments, dual-task paradigms wherein subjects performed a primary sentence recognition task simultaneously with a secondary task were used. Because the dual-task interference is dependent on factors such as task demand and degree to which overlapping resources are required (Hazeltine et al. 2006; Wickens 2008), two different secondary tasks, one being simpler and one being more complex, were used in the study to examine the effect of secondary task on the shape of psychometric function. Because objective and subjective measures may assess different aspects of listening effort (e.g., Fraser et al. 2010; Zekveld et al. 2010), the current study also characterized the psychometric function of self-reported listening effort.

EXPERIMENT 1

The purpose of Experiment 1 was to characterize the pattern of task performance across SNRs in dual-task listening effort measures for YNH listeners.

Materials and Methods

Participants—In total 26 younger adults were recruited. Most of them were college students. Two participants were unable to complete the study due to time conflict of their second laboratory visit (see below for experiment procedures). For the 24 YNH (12 males and 12 females) participants who completed the study, their ages ranged from 19 to 30 years with a mean of 23.7 years ($SD = 3.6$). The participants had pure-tone thresholds better than 25 dB HL at 0.5, 1, 2, and 4 kHz (ANSI 2010). All participants are native speakers of English. The sample size was determined based on a pilot study, which indicated that a sample size of 24 were needed in order to detect the effect of SNR on task performance (assuming $\alpha = 0.05$ and power = 0.8).

Dual-task paradigm—Two dual-task paradigms that included different secondary tasks were used. In the dual-task measure the participants performed a speech recognition task simultaneously with either an easy or hard secondary task.

Primary speech recognition task: The Hearing in Noise Test (HINT) (Nilsson et al. 1994) was used as the speech material. In order to ensure that the dual-task measure was conducted across a wide range of speech intelligibility for each participant, individualized SNRs were used. Specifically, before the dual-task measure, an individual SNR-50 at which the participant could understand 50% of speech was determined using an adaptive SNR procedure. During the SNR-50 measure, the participants listened to the HINT sentences and repeated as much as possible. The speech level was fixed at 65 dB SPL. The level of the noise, which was the speech-shaped noise of the HINT, was adjusted adaptively depending on the participant's responses using the one-down, one-up procedure in 2 dB steps. The correct response of each sentence was based on the repetition of the whole sentence, with minor exceptions such as a/the and is/was. Twenty HINT sentences were used and the SNRs of the last 16 sentences were averaged (Nilsson et al. 1994). This averaged SNR minus 2 dB was defined as an individual's SNR-50. According to the pilot study, at this SNR-50 a listener's speech recognition performance would be close to 50% correct if the scoring of the HINT was based on words (which was the scoring method used in the dual-task measures).

For a given participant, 11 SNRs ranging from -10 to $+10$ dB in 2-dB steps relative to this participant's SNR-50 were created and used in the speech recognition task of the dual-task measure. The speech presentation level was fixed at 65 dB SPL for all 11 SNR conditions. The HINT sentences (different sentences from those used in the SNR-50 measure) and speech-shaped noise were used. For each of the 11 SNR conditions, 20 trials (20 HINT sentences) were conducted. In each trial, the noise was presented 1 sec before the onset the sentence and ended approximately 1 sec after the offset of the sentence. The trials that had the same SNR were administered in one block. The order of SNR block was randomized. The participants were asked to repeat as much of each sentence as possible. The

participants' performance (i.e., the HINT score) at a given SNR was quantified by dividing the number of words the participants repeated correctly by the total number of words in the 20 sentences.

Secondary task: The visual stimuli of the Stroop test (Stroop 1935) were used in the secondary task. During testing, visual stimuli of color words displayed in different font colors were shown in the middle of a computer monitor. Four color words and font colors were used: red, blue, green, and yellow. The combination of color word and font color was randomized, but the color word was always inconsistent with the font color. Below the stimulus word, the computer monitor showed four boxes containing "RED", "BLUE", "GREEN", and "YELLOW," respectively, to represent the four virtual response buttons. The font color of the words in the virtual button box was black.

Using the same visual stimuli, two tasks were created. For the easy task, listeners were asked to press the space bar on the keyboard as quickly as possible after stimulus word presentation, regardless of the word and the font color. In other words, the easy task was a simple visual reaction-time task. When the space bar was pressed, all four virtual buttons on the screen were highlighted to indicate the response.

In the hard task, the Stroop test paradigm was used. In particular, listeners were asked to respond to the font color, instead of the word, by pressing a keyboard button assigned to a given color as quickly as possible (i.e., the incongruent condition of the Stroop test). The keyboard buttons "D", "C", "M", and "K" were assigned to font color red, blue, green, and yellow, respectively. To assist the participants in determining which keyboard button to press during the testing, the relative position of the four virtual buttons on the screen was arranged to spatially map that of the four keyboard buttons. When a given button was pressed, the corresponding virtual button on the screen was highlighted to indicate the response. Because the participants needed to inhibit the semantic meaning of the stimulus word and determine which button to push in the hard task, this task was more demanding and would interfere more with the speech recognition task than the easy task.

For both the easy and hard tasks, one stimulus word was presented with one HINT sentence in each trial. Because the total processing load for sentence understanding reaches a maximum at the end of the sentence (Winn et al. 2015), the stimulus word was presented at a random time during the second half of each HINT sentence presentation. The RT from stimulus word onset until a keyboard button was pressed quantified the performance of the secondary task.

Subjective effort rating—The participants were asked to rate their perceived listening effort after listening to 20 HINT sentences in a given SNR block of the dual-task measure. The participants answered the question "how hard were you trying to understand the speech" using a 21-point scale ranging from 0, representing "not at all," to 100, representing "very, very hard." The question and the scale were adapted from the "effort" question of the National Aeronautics and Space Administration Task Load Index questionnaire (Hart & Staveland 1988). Likely due to the wide SNR range used in the experiment, ceiling effect was observed on some subjects in the pilot study. In order to minimize the ceiling (and floor)

effect, the participants were allowed to use larger numbers than 100 and negative numbers to rate listening effort (Hällgren et al. 2005). For this reason, two arrowheads were placed at the two ends of the 21-point scale, one on each end, pointing away from the scale. Because the pilot study indicated that listeners did not report the highest effort at the poorest SNR, the participants were not trained to use the scale; i.e., they were not provided with samples of the most favorable and unfavorable SNRs to anchor the end points of the scale.

Procedures—The study was approved by the Institutional Review Board of the University of Iowa. After agreeing to participate in the study and signing the consent form, the participants' pure tone thresholds were measured, followed by the SNR-50 measure. Afterward, the dual-task tests were conducted. During the testing, the participants were instructed to repeat as much of each sentence as possible and respond to the visual stimulus as quickly (and accurately) as possible. The participants were asked to give priority to the speech recognition task, i.e., they should always try to maximize their speech recognition performance. The participants were also asked not to let the repetition of sentences affect their response speed to the visual stimuli of the secondary task, i.e., they should not repeat the sentence until they pushed the keyboard button. Prior to the experiment, a tutorial and practice session was given to familiarize participants with the tasks. For each secondary task, at least 60 trials were given in the practice. The dual-task experiment was conducted in 22 conditions (2 secondary tasks x 11 SNRs). Due to the limited length of the HINT, the HINT sentences were used twice in the experiment. Therefore, the experiment was completed in two laboratory visits: one for the easy task and one for the hard task. The interval between the two visits was at least one month in order to minimize the learning effect. The order of easy/hard task was randomized across the participants. Momentary compensation was provided to the participants at the completion of the study.

The experiment was conducted in a sound treated booth. All auditory stimuli (HINT sentences and noise) were presented via earphones. The auditory stimuli were generated by a computer and an M-Audio (Cumberland, Rhode Island) ProFire 610 sound interface, routed to a Grason-Stadler (Eden Prairie, Minnesota) GSI-61 audiometer, an Alesis (Cumberland, Rhode Island) DEQ830 digital equalizer, a Samson (Hauppauge, New York) Servo 120 amplifier, and then presented through a pair of Sennheiser (Wedemark, Germany) IE8 insert earphones. The output of the earphones was calibrated in a G.R.A.S. (Holte, Denmark) IEC 711 RA0045 Ear Simulator. The visual stimuli were presented using a 19 in. computer monitor, which was placed in front of the participants. The psychological testing software E-prime 2.0 (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania) was used to present auditory and visual stimuli, collect participants' responses, and measure the RT.

Data analysis—Data were first processed before analysis. For the HINT score (in percent), a logit transformation was conducted to linearize the relationship and homogenize the variance (logit-transformed score = $\log((\text{HINT score} + 1) / (101 - \text{HINT score}))$). For the easy secondary task, because the distribution of the RT across 20 trials at a given SNR was skewed, the median of the 20 RTs served as the RT of that SNR condition and was used in analysis. For the hard task (the Stroop test), the response accuracy was first examined. The overall accuracy across all conditions and participants was 95.3% (SD = 5%). The

Friedman repeated measures analysis of variance on ranks indicated that SNR did not have an effect on response accuracy ($p = 0.78$), suggesting that Stroop accuracy did not vary with speech intelligibility. The difference in Stroop accuracy between the participants' first and last conditions was also found to be non-significant ($p = 0.13$), indicating that the learning effect of the Stroop test was minimal. Because the Stroop accuracy was high and was stable across SNRs and time, the median RT across all 20 trials at a given SNR, regardless of accuracy, served as the RT of that SNR and was used in analysis. The distribution of the median RT across all participants, 11 SNRs, and two secondary tasks was then examined. Because the distribution was right-skewed, the RT was log-transformed before analysis (log-transformed RT = $\log(\text{RT})$). For the subjective effort rating, the distribution was first examined. Among all ratings, 12 ratings (2.2%) had values larger than 100 and the highest rating was 200. No rating was smaller than 0. Because some participants used larger numbers than 100 to rate listening effort, the subjective effort rating was linearly transformed so that the scale of the rating was the same across all participants. In particular, for a given participant, the subjective effort rating was linearly rescaled such that the maximum and minimum ratings across all 22 conditions (11 SNRs x 2 secondary tasks) were 100 and 0, respectively. Because the distribution of the rescaled subjective listening effort rating across all participants and test conditions was normal, no further transformation was conducted.

To characterize the psychometric function (i.e., the trend of performance across SNRs) of dual-task listening effort measures, a linear mixed model was used to analyze the repeated measure data. In particular, the model fits polynomial (linear, quadratic, and cubic) terms of SNR to the HINT score (logit-transformed), RT (log-transformed), or subjective effort rating data (rescaled) for each participant. The SNR quadratic and cubic terms were rescaled using quadratic term = $(\text{SNR}/5)^2$ and cubic term = $(\text{SNR}/5)^3$ to aid in convergence. In polynomial models, the linear term reflects an overall slope of the function; the quadratic term reflects the shape of the primary inflection point of the curve (positive and negative coefficients reflect concave-upward and -downward curves, respectively), and the cubic term reflects the steepness of the secondary inflection point. Fixed effects considered in the model were a secondary task term (easy/hard), three SNR polynomial terms (linear, quadratic, and cubic), and all two-way interactions. A random intercept, random SNR linear, quadratic and cubic terms were also included in the model. The random-effect term was then removed one by one to identify the model that had minimum Akaike Information Criterion (AIC). For the HINT score and RT, the model that included the random effect of intercept and SNR linear term had minimum AIC and was selected; for subjective effort rating, the model that included the random effect of intercept and SNR linear and quadratic terms had minimum AIC and was selected. The effects of other variables were estimated as fixed effects only. After the model was settled with respect to the random effect, the fixed-effect terms were examined. When an interaction term was not significant, it was excluded from the final model. The SNR cubic term was not included in the final model if it was not significant. The analysis was conducted using the Statistical Analysis Software (SAS Institute, Cary, North Carolina).

Results

Speech recognition performance—Figure 1A shows the mean HINT scores across all participants for each secondary task as a function of SNR. The two sigmoidal curves are almost overlapped. The analysis first indicated that none of the interactions was significant. The results further showed that secondary task did not have a significant effect on HINT score ($\beta = 0.05$, $F_{1, 23} = 0.80$, $p = 0.38$). In contrast, SNR linear ($\beta = 0.65$, $F_{1, 500} = 2599.1$, $p < 0.0001$), quadratic ($\beta = -0.05$, $F_{1, 500} = 6.57$, $p = 0.010$), and cubic terms ($\beta = -0.26$, $F_{1, 500} = 179.3$, $p < 0.0001$) were all significant. These results indicated that HINT score increased as SNR increased and that this trend was similar for both the easy and hard dual-task tests.

Secondary task performance—Figure 1B shows the RT (the median RT of a given SNR) averaged across all participants for each secondary task at each SNR. Longer RT represents poorer performance. For both tasks, the RT curves had a peaked shape. The RT increased as SNR changed from favorable (i.e., +8 and +10 dB) to intermediate SNRs (0 and -2 dB), and then decreased as SNRs moved from intermediate to unfavorable SNRs (-8 and -10 dB). To better compare the shape of the curves, in Figure 1C the y-axis of the two RT curves shown in Figure 1B was rescaled so that the two curves have similar peak heights in the figure. Figure 1C shows that the two curves are very similar in shape and the curves are fairly symmetrical around the peaks.

The analysis revealed that the interactions and the SNR cubic term were not significant. The effect of secondary task was significant ($\beta = -0.93$, $F_{1, 23} = 2585.0$, $p < 0.0001$). The SNR linear term was significant ($\beta = -0.005$, $F_{1, 501} = 6.14$, $p = 0.014$). The negative coefficient indicated that the right side of the curve is generally lower than the left side (Figure 1C). The SNR quadratic term was also significant ($\beta = -0.05$, $F_{1, 501} = 51.7$, $p < 0.0001$). The negative coefficient confirmed the peaked shape of the curves.

Subjective effort rating—After listening to 20 HINT sentences at a given SNR block, the participants rated their listening effort. Higher ratings represented more listening effort. Figure 2D shows the rescaled subjective effort rating averaged across all participants as a function of SNR. Essentially, these curves also display a peaked shape. Compared to the RT curves shown in Figure 1C, the subjective effort rating curves are less symmetric around the peaks.

Mixed-effects analysis first revealed that the interactions and the SNR cubic terms were not significant. The effect of secondary task was significant ($\beta = 4.43$, $F_{1, 23} = 5.65$, $p = 0.026$), indicating that the participants reported that they tried harder to understand speech in the easy dual-task measure than in the hard dual-task measure. The results further indicated that, while the SNR linear term was not significant ($\beta = -0.78$, $F_{1, 501} = 2.49$, $p = 0.12$), the quadratic term was ($\beta = -6.24$, $F_{1, 501} = 25.5$, $p < 0.0001$).

Discussion

The results of Experiment 1 showed that the RT of the secondary task in the dual-task measure had a non-linear trend over SNRs; that is, as SNR decreased, the RT first increased

and then decreased (Figures 1B and 1C). More specifically, at favorable SNRs (speech intelligibility close to 100%), the RT was short, suggesting that speech recognition was easy and did not require much top-down processing. As SNR decreased, the RT increased, indicating that the participants used more top-down working memory processing to process the degraded speech signals. This was consistent with the ELU model (Rönnerberg et al. 2008). The RT reached its peak at -2 dB or 0 dB relative to SNR-50 (speech intelligibility = 30% to 50%). As the SNR kept decreasing, however, speech recognition performance became poorer while the RT became shorter. The shorter RT suggested that cognitive processing was shifted from the speech recognition task to the secondary task. The peaked shape of RT curve was consistent with Granholm et al (1996) and the second experiment of Zekveld and Kramer (2014), suggesting that the participants might experience cognitive overload at the unfavorable SNRs.

Recall that the hard secondary task (Stroop test) required the participants to inhibit the semantic meaning of the stimulus word and determine which button to push, while the easy task was a simple visual reaction-time task. Therefore, it is not surprising that the RT of the hard task (0.85 sec, averaged across SNRs) was longer than that of the easy task (0.35 sec). Despite the large difference between the two secondary tasks, the RT curves of the easy and hard tasks had similar shapes, as indicated by the non-significant interaction between secondary task and SNR polynomial terms. The implication of this finding will be further discussed in the General Discussion section at the end of this paper.

Similar to the RT curve, the curve of subjective effort rating had a non-linear trend (Figure 1D). This result is not in line with the first experiment of Zekveld and Kramer (2014), which found that both pupil response and subjective listening effort rating increased linearly as SNR and speech intelligibility decreased.

Of note, the participants reported higher listening effort in the easy than the hard dual-task measures. One speculation on this finding is that the self-reported rating reflected effort allocation between the primary and secondary tasks. Specifically, because the visual-reaction time task was less demanding than the Stroop test, the participants were able to exert more effort to the speech recognition task in the easy than the hard dual-task measure. As a result, when the participants were asked to answer the question “how hard were you trying to understand the speech,” they reported that they tried harder in the easy dual-task measure.

EXPERIMENT 2

The purpose of Experiment 2 was to replicate Experiment 1 using older listeners with hearing impairment (OHI).

Materials and Methods

In total 24 OHI (10 males and 14 females) participants were recruited and completed the study. Their ages ranged from 56 to 83 years with a mean of 69.9 years ($SD = 5.8$). The participants were eligible for inclusion in this study if their hearing loss met the following criteria: (1) postlingual bilateral downward-sloping sensorineural hearing loss (air-bone gap < 10 dB); (2) hearing thresholds no better than 20 dB HL at 500 Hz and no worse than 85

dB HL at 3 kHz (ANSI 2010); and (3) hearing symmetry within 15 dB for all test frequencies. The mean pure tone thresholds are shown in Figure 2. All participants were native speakers of English.

The stimuli, test conditions, equipment and procedures were identical to those used in Experiment 1. Experiment 2 differed from Experiment 1 in that the auditory stimuli (HINT sentences and noise) were spectrally shaped and linearly amplified before being routed to the earphones. The purpose of amplifying the stimuli was to ensure that speech intelligibility could approach 100% at favorable SNRs for all participants. The individual frequency shaping and amplification were based on each participants' audiometric thresholds and the NAL-NL2 formula (Keidser et al. 2011). Specifically, from an Audioscan Verifit hearing aid analyzer, the NAL-NL2 targets of real ear aided responses (REAR) to a 65-dB SPL speech input (the "carrot passage") from 0.25 to 6 kHz for each participant were obtained. The REAR targets were used to configure the filter and gain settings of the DEQ830 multi-channel equalizer, one channel for each ear, to shape the one-third octave band spectra of the input signals such that, for the 65-dB "carrot passage" input, the outputs of the earphones met an individual's NAL-NL2 REAR targets within ± 3 dB across 0.25 to 6 kHz. Using the amplified auditory stimuli, the participant's SNR-50 was measured and the 11 SNRs (-10 to $+10$ dB relative to SNR-50, 2-dB steps) were created. Before the testing, the participants were asked about their loudness perception of the stimuli. All participants reported that the sound level was appropriate.

Identical to Experiment 1, the response accuracy of the hard secondary task (the Stroop test) was examined before data analysis. The overall accuracy across all conditions and participants was high (97.9%). The accuracy did not vary with SNR ($p = 0.68$) and was not different between the participants' first and last condition ($p = 0.76$). Therefore, for both the easy and hard tasks, the median RT across all 20 trials at a given SNR served as the RT of that SNR condition. For the subjective effort rating, the distribution was first examined. Among all ratings, 34 ratings (6.4%) had values larger than 100 and no rating was smaller than 0. The highest rating was 200. Identical to Experiment 1, the subjective effort rating was linearly transformed so that the scale of the rating was the same across all participants. Mixed-effects analysis was then used to determine the effect of secondary task and SNR polynomial terms on HINT score (logit-transformed), RT (log-transformed), or subjective effort rating (rescaled). The fixed and random effects that were included in the models were identical to those in Experiment 1.

Results

Speech recognition performance—The four panels in Figure 3 shows HINT score, RT, rescaled RT, and rescaled subjective effort rating as a function of SNR. For the HINT score (Figure 3A), analysis indicated that none of the interactions was significant. The effect of secondary task was not significant either ($\beta = 0.026$, $F_{1, 23} = 0.21$, $p = 0.65$). In contrast, the SNR linear ($\beta = 0.60$, $F_{1, 500} = 1889.9$, $p < 0.0001$), quadratic ($\beta = -0.13$, $F_{1, 500} = 39.5$, $p < 0.0001$), and cubic terms ($\beta = -0.22$, $F_{1, 500} = 133.3$, $p < 0.0001$) were all significant.

Secondary task performance—Similar to the YNH participants in Experiment 1, the RT curves of the OHI participants were peak shaped (Figures 3B and 3C). The peaks were located at -2 dB relative to SNR-50 for both the easy and hard tasks. Analysis results first revealed that none of the interactions was significant, nor was the SNR cubic term. The results further indicated that the easy task's RT was significantly shorter than that of the hard task ($\beta = -0.92$, $F_{1, 23} = 1555.1$, $p < 0.0001$). While the SNR linear term was not significant ($\beta = 0.002$, $F_{1, 501} = 2.07$, $p = 0.15$), the quadratic term was ($\beta = -0.073$, $F_{1, 501} = 78.7$, $p < 0.0001$),

Subjective effort rating—Although the OHI participants' subjective effort rating curve had a peaked shape in the hard dual-task measure, the curve in the easy dual-task measure was more like a reversed sigmoidal shape (Figure 3D). Mixed-effects analysis showed that subjective listening effort rating was higher in the easy than the hard dual-task measure ($\beta = 7.55$, $F_{1, 23} = 17.9$, $p = 0.0003$). The SNR linear ($\beta = -4.54$, $F_{1, 498} = 45.5$, $p < 0.0001$), quadratic ($\beta = -7.04$, $F_{1, 498} = 37.6$, $p < 0.0001$), cubic terms ($\beta = 4.18$, $F_{1, 498} = 10.8$, $p = 0.001$) were also significant. The results further indicated that the interaction between secondary task and SNR linear term ($\beta = 2.25$, $F_{1, 498} = 9.5$, $p = 0.002$) and the interaction between secondary task and SNR cubic term ($\beta = -4.48$, $F_{1, 498} = 14.5$, $p = 0.0002$) were significant.

Discussion

Generally, the results of the OHI participants in this experiment were consistent with the findings of the YNH subjects in Experiment 1. For both the easy and hard secondary tasks, the RT curve had a non-linear trend such that the RT initially increased as speech recognition became more difficult and decreased when the HINT score was lower than 30%. For the subjective effort rating, the participants reported more listening effort in the easy than the hard dual-task measure. However, inconsistent with Experiment 1, the significant interaction between secondary task and SNR polynomial terms indicated that the trend of subjective effort rating across SNRs was different between the easy and the hard dual-task measure. This difference mainly resulted from the large discrepancy in effort rating at the unfavorable SNRs (Figure 3D). It is unclear why the OHI participants reported high effort at the unfavorable SNRs in the easy dual-task measure but not in the hard dual-task measure, and why this pattern was not observed in the YNH participants of Experiment 1.

EXPERIMENT 3

In both Experiments 1 and 2, 20 dual-task trials that had the same SNR were administered in one block. Because the SNR was fixed across 20 trials, the participants could obtain a general idea about the test difficulty and their speech recognition performance level after the first few trials. As a result, at unfavorable SNRs the participants might easily decide to give up on the listening task in the rest of trials, resulting in shorter RTs. If the SNR was varied from trial to trial, the RT curve might have a different shape. Zekveld and Kramer (2014) have speculated that SNR presentation order (blocked vs. non-blocked) may explain why their two experiments generated inconsistent results regarding the trend of pupil response across SNRs. To investigate if the peaked shape of the RT curve was specific to the blocked

SNR design used in Experiments 1 and 2, Experiment 3 characterized the RT curve in dual-task paradigms, wherein the SNR was varied from trial to trial, for YNH listeners.

Materials and Methods

In total 25 YNH (12 males and 13 females) participants were recruited and completed the study. Most of the participants were college students and their ages ranged from 19 to 30 years with a mean of 21.2 years ($SD = 2.7$). The participants had pure-tone thresholds better than 25 dB HL at 0.5, 1, 2, and 4 kHz (ANSI 2010). All participants were native speakers of English.

The stimuli, dual-task paradigms, test SNRs, equipment, procedures, and data transformation and analysis were identical to those used in Experiment 1. Experiment 3 differed from Experiment 1 in that, for a given secondary task, the presentation SNR was randomized across the 220 HINT trials (20 sentences \times 11 SNRs). Because SNR presentation order was randomized, the participants were not asked to report their perceived listening effort.

Results

Figure 4 shows the results. For the HINT score (Figure 4A), analysis indicated that secondary task did not have a significant effect of HINT score ($\beta = 0.014$, $F_{1, 24} = 0.09$, $p = 0.77$), nor did the SNR quadratic term ($\beta = -0.016$, $F_{1, 521} = 0.89$, $p = 0.35$). The effect of SNR linear term ($\beta = 0.64$, $F_{1, 521} = 3738.3$, $p < 0.0001$) and cubic term ($\beta = -0.25$, $F_{1, 521} = 243.3$, $p < 0.0001$) was significant.

The RT curves of both the easy and hard tasks were peak shaped (Figures 4B and 4C). The curve peaks of the easy and hard tasks were at -2 and 0 dB relative to SNR-50, respectively. Mixed-effects analysis revealed that none of the interactions was significant, nor was the SNR cubic term. In contrast, the secondary task term ($\beta = -1.04$, $F_{1, 24} = 5018.8$, $p < 0.0001$) and SNR linear ($\beta = -0.003$, $F_{1, 522} = 3.91$, $p = 0.04$) and quadratic terms ($\beta = -0.032$, $F_{1, 522} = 37.6$, $p < 0.0001$) were all significant. Similar to Experiment 1, the linear trend was negative (i.e., the right side of the curve was lower than the left side).

Individual difference—Averaged across the participants, the RT curve had a peaked shape and the peak was located at -2 to 0 dB relative to SNR-50 (Figure 4C). At the individual level, however, the shape of the RT curve varied considerably. To illustrate this point, Figure 5 shows the hard task RT curve of four participants in Experiment 3. In this figure, the top three curves have a peaked shape, but the width and location of the peak varies. The curve shown at the bottom of the figure has a reversed sigmoidal shape instead of a peaked shape. Although Figure 6 shows the results only from Experiment 3, large individual difference has been observed across all three experiments of the current study.

Comparison across three experiments—To examine if SNR presentation order (blocked vs. non-blocked) had an effect on the trend of RT, analysis on the data collected from Experiments 1 and 3 was conducted. Because it is also of interest to compare the RT curve of YNH and OHI participants, the data of Experiment 2 were included in the analysis.

Mixed-effects analysis was performed to investigate the effect of secondary task (easy/hard), SNR polynomial terms, and experiment (between-subject variable, Experiments 1/2/3) on RT (log-transformed). The model included the random effect of intercept and SNR linear term. The effects of other variables were estimated as fixed effects only. Figure 6 summarizes the RT curves of each experiment and each secondary task.

The results revealed that the main effects of secondary task ($F_{1, 72} = 3144.8, p < 0.0001$), experiment ($F_{2, 70} = 19.0, p < 0.0001$), and SNR quadratic term ($\beta = -0.023, F_{1, 1525} = 166.3, p < 0.0001$) were all significant, while the SNR linear term ($\beta = -0.003, F_{1, 1525} = 3.35, p = 0.067$) was not. The interaction between experiment and SNR linear term ($F_{2, 1525} = 5.25, p = 0.005$) and the interaction between experiment and SNR quadratic term ($F_{2, 1525} = 9.56, p < 0.0001$) were also significant. The post-hoc comparison showed that the SNR linear and quadratic terms of Experiment 1 did not significantly differ from those of Experiment 3 ($p = 0.45$ and $p = 0.12$, respectively). In contrast, the SNR linear and quadratic terms were significantly different between Experiments 1 and 2 ($p = 0.002$ and $p = 0.006$, respectively) and between Experiments 2 and 3 ($p = 0.017$ and $p < 0.0001$, respectively).

Discussion

The RT linear and quadratic terms did not differ between Experiments 1 and 3, suggesting that RT curve shape was not affected by SNR presentation order (blocked vs. non-blocked). Note that this result does not necessarily exclude the possibility that the participants actively quit listening at the unfavorable SNRs: the participants might quickly decide to give up on the listening task right after the onset of noise, which was presented 1 sec before the speech.

In contrast, the RT curves of YNH listeners (Experiments 1 and 3) and OHI listeners (Experiment 2) had different shapes. The difference in the SNR quadratic term was because the RT curve of the YNH participants was flatter than that of the OHI participants (Figure 6). This may reflect that OHI listeners exert more effort on speech understanding than YNH listeners (Desjardins & Doherty 2013; Degeest et al. 2015). The difference in the linear trend was because the RT curve showed a negative trend in Experiments 1 and 3 (YNH) but not in Experiment 2 (OHI). This difference indicated that YNH participants' RTs at unfavorable SNRs were longer than that at favorable SNRs, while OHI listeners' RTs were similar for both the unfavorable and favorable SNRs. The relatively short RT of OHI listeners at unfavorable SNRs may suggest that these listeners experienced more cognitive overload than YNH participants. This speculation was consistent with the study by Petersen et al (2015), which found that alpha power breakdown is more likely to occur for listeners with more severe hearing loss in the most difficult condition.

GENERAL DISCUSSION AND CONCLUSIONS

The three experiments of the current study examined the task performance of dual-task listening effort measures across a wide range of SNR and speech intelligibility. The results suggested that RT had a non-linear trend across SNRs: RT was the longest at -2 dB or 0 dB relative to SNR-50 and was shorter when speech intelligibility was better than 50% or poorer than 30%. This pattern was observed for both YNH and OHI participants and was not affected by either the type of secondary task (easy or hard) or SNR presentation order

(blocked or non-blocked). The result showing that RT reached its peak when speech intelligibility was between 30% and 50% was in line with the second experiment of Zekveld and Kramer (2014), which found that pupil size was the largest when speech intelligibility was approximately 50%.

Why was the RT shorter at the unfavorable SNRs than the intermediate SNRs? As mentioned, this can be explained by the tendency of actively giving up listening in cognitive overload situations. Because the current study used dual-task paradigms, the peaked shape of the RT curve can also be explained by the adaptive gain theory (Aston-Jones & Cohen 2005). In particular, this theory tries to explain the neurophysiological mechanism of the trade-off between an animal's exploitative behavior (optimizing the performance of the current task) and exploratory behavior (searching for alternative sources of reward). The adaptive gain theory assumes that the trade-off between these two behaviors is driven by on-line assessments of task-relevant utility; that is, the costs and benefits associated with the task. It is likely that, in the dual-task measures of the current study, the utility of the primary speech recognition task was high at the favorable and intermediate SNRs and the participants expended effort on this task to optimize the performance. As the utility in the speech recognition task waned at very unfavorable SNRs, the participants disengaged themselves from the listening task and exerted more effort on the secondary task to pursue reward.

If the participants disengaged themselves from the listening task, can they “work harder” to improve their speech recognition performance? According to the ELU model, explicit and deliberate working memory top-down processes are invoked when the speech information input is degraded. Therefore, the longer RTs at intermediate SNRs suggest that the participants dedicated more working memory processes at these SNRs than at other SNRs. That is, even when their speech recognition performance was lower than 100% and there was room for improvement, at SNRs other than the intermediate SNRs (including more favorable and unfavorable SNRs) the participants did not allocate all available working memory resources to speech processing. It is unclear whether it is possible for listeners to deliberately dedicate more working memory processes to the task, and if in doing so can improve their speech understanding. Clarifying these issues may advance our understanding about the cognitive mechanism of speech listening in adverse conditions.

Recall that the motivation of the current study was to determine the optimal speech intelligibility level for dual-task listening effort measures. The study results indicate that, due to the RT nonlinear trend across SNRs, the dual-task paradigm should be used cautiously as a measure of listening effort. For example, if a hearing aid technology can improve SNR (e.g., directional microphones), dual-task measures could demonstrate that this technology improves, decreases, or has no effect on secondary task performance, depending on the test SNR and speech intelligibility level. If the change in secondary task performance is taken as an index of the change in listening effort, the result of dual-task measures could show that this hearing aid technology improves speech intelligibility while increasing listening effort. In order to avoid this paradoxical result, it is suggested that a dual-task listening effort measure is conducted at speech intelligibility level higher than 50%. If speech signal is highly degraded and the intelligibility is lower than 30%,

individuals may experience cognitive overload and/or disengage the listening task. As a result, data interpretation will be more complex.

Although the peaked shape of the RT curve was consistently observed across the easy and hard secondary tasks used in the current study, this result may not generalize to all dual-task listening effort measures. For example, Pals et al (2013) used dual-task paradigms to measure the effect of spectral resolution (vocoder simulation) on listening effort for YNH listeners. Two different secondary tasks were used: a rhyme-judgment task and a mental rotation task. The results showed that, as the number of spectral channels decreased from 24 to 2 channels, speech intelligibility decreased from 100% to approximately 15% and RT of both secondary tasks increased monotonically. The RT was the longest in the lowest intelligibility (2-channel) condition. It is unclear why in Pals et al (2013) the RT trend across channel number did not show a peaked shape as the currently study. Possible explanations, which include the difference in speech signal degradation (vocoded speech vs. speech in noise) and secondary task, should be explored in future work.

It was observed that the shape of individual RT curve varied considerably across participants (Figure 5). This variation can be regarded as a limitation of the study because the study result (i.e., the peak-shaped psychometric function) does not hold for all individuals. This individual difference, however, may reflect how people cope with adverse listening conditions. In particular, individuals who have peak-shaped RT curves may be more likely to experience cognitive overload and/or give up listening than those who have reversed-sigmoidal RT curves. Therefore, the former listeners may tend to use maladaptive strategies (e.g., pretending to understand the conversation) to avoid unpleasant situations, while the latter listeners may be more likely to use adaptive strategies (e.g., asking the talker to repeat) to improve communication (Demorest & Erdman 1987). More research is needed to investigate these speculations.

Acknowledgments

Experiments 1 and 2 were supported by the New Century Scholars Research Grant from American Speech-Language-Hearing Foundation. Experiment 3 was supported by NIH/NIDCD R03DC012551 and R01-DC012769. The pilot study of this project was supported by a research grant from Siemens Hearing Instruments. The authors also thank Drs. Bob McMurray and Jacob Oleson for their valuable comments and suggestions on an early version of this paper.

References

- ANSI. Specification for Audiometers (ANSI S3.6). New York: American national standards institute; 2010.
- Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci*. 2005; 28:403–450. [PubMed: 16022602]
- Cabestrero R, Crespo A, Quirós P. Pupillary dilation as an index of task demands. *Percept Motor Skill*. 2009; 109:664–678.
- Degeest S, Keppler H, Corthals P. The effect of age on listening effort. *J Speech Lang Hear Res*. 2015; 58:1592–1600. [PubMed: 26161899]
- Demorest ME, Erdman SA. Development of the communication profile for the hearing impaired. *J Speech Hear Disord*. 1987; 52:129–143. [PubMed: 3573744]
- Desjardins JL, Doherty KA. Age-related changes in listening effort for various types of masker noises. *Ear Hear*. 2013; 34:261–272. [PubMed: 23095723]

- Desjardins JL, Doherty KA. The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear.* 2014; 35:600–610. [PubMed: 24622352]
- Downs DW. Effects of hearing and use on speech discrimination and listening effort. *J Speech Hear Disord.* 1982; 47:189–193. [PubMed: 7176597]
- Fraser S, Gagné JP, Alepins M, et al. Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J Speech Lang Hear Res.* 2010; 53:18–33. [PubMed: 19635945]
- Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. *Br J Audiol.* 1990; 24:63–68. [PubMed: 2317602]
- Gosselin PA, Gagné JP. Use of a dual-task paradigm to measure listening effort. *Canadian J Speech Lang Patho Audiol.* 2010; 34:43–51.
- Gosselin PA, Gagné JP. Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res.* 2011; 54:944–958. [PubMed: 21060138]
- Granhölm E, Asarnow RF, Sarkin AJ, et al. Pupillary responses index cognitive resource limitations. *Psychophysiology.* 1996; 33:457–461. [PubMed: 8753946]
- Hällgren M, Larsby B, Lyxell B, et al. Speech understanding in quiet and noise, with and without hearing aids. *Int J Audiol.* 2005; 44:574–583. [PubMed: 16315448]
- Hart, SG.; Staveland, LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, PA.; Meshkati, N., editors. *Human Mental Workload.* Amsterdam, The Netherlands: North Holland Press; 1988. p. 139-183.
- Hazeltine E, Ruthruff E, Remington RW. The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychol.* 2006; 52:291–345.
- Hick CB, Tharpe AM. Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res.* 2002; 45:573–584. [PubMed: 12069009]
- Hornsby BWY. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear.* 2013; 34:523–534. [PubMed: 23426091]
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007; 18:590–603. [PubMed: 18236646]
- Kahneman, D. *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall; 1973.
- Keidser G, Dillon HR, Flax M, et al. The NAL-NL2 prescription procedure. *Audiol Res.* 2011; 1:88–90.
- Kiessling J, Pichora-Fuller MK, Gatehouse S, et al. Candidature for and delivery of audiological services: special needs of older people. *Int J Audiol.* 2003; 42:2, S92–101. [PubMed: 12564509]
- McGarrigle R, Munro KJ, Dawes P, et al. Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group ‘white paper’. *Int J Audiol.* 2014; 53:433–445. [PubMed: 24673660]
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* 1994; 95:1085–1099. [PubMed: 8132902]
- Pals C, Sarampalis A, Başkent D. Listening effort with cochlear implant simulations. *J Speech Lang Hear Res.* 2013; 56:1075–1084. [PubMed: 23275424]
- Peavler, WS. Individual differences in pupil size and performance. In: Janisse, MP., editor. *Pupillary dynamics and behavior.* Springer; 1974. p. 159-175.
- Petersen EB, Wöstmann M, Obleser J, et al. Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front Psychol.* 2015; 6:1–11. [PubMed: 25688217]
- Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiological rehabilitation. *Trends Amplif.* 2006; 10:29–59. [PubMed: 16528429]
- Picou EM, Ricketts TA. The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear Hear.* 2014; 35:611–622. [PubMed: 24992491]
- Picou EM, Ricketts TA, Hornsby BW. How hearing aids, background noise, and visual cues influence objective listening effort. *Ear Hear.* 2013; 34:e52–e64. [PubMed: 23416751]

- Poock GK. Information processing vs pupil diameter. *Percept Motor Skill*. 1973; 37:1000–1002.
- Rönnerberg J. Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int J Audiol*. 2003; 42:S68–S76. [PubMed: 12918612]
- Rönnerberg J, Lunner T, Zekveld AA, et al. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci*. 2013; 7:1–17. [PubMed: 23420631]
- Rönnerberg J, Rudner M, Foo C, et al. Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol*. 2008; 47:S99–S105. [PubMed: 19012117]
- Sander MC, Werkle-Bergner M, Lindenberger U. Amplitude modulations and inter-trial phase stability of alpha-oscillations differentially reflect working memory constraints across the lifespan. *Neuroimage*. 2012; 59:646–654. [PubMed: 21763439]
- Sarampalis A, Kalluri S, Edwards B, et al. Objective measures of listening effort: effects of background noise and noise reduction. *J Speech Lang Hear Res*. 2009; 52:1230–1240. [PubMed: 19380604]
- Stroop J. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935; 18:643–662.
- Wickens CD. Multiple resources and mental workload. *Hum Factors*. 2008; 50:449–455. [PubMed: 18689052]
- Winn MB, Edwards JR, Litovsky RY. The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear Hear*. 2015; 36:e153–e165. [PubMed: 25654299]
- Worrall, LE.; Hickson, L. Communication disability in aging from prevention to intervention. New York: Thomson, Delmar Learning; 2003.
- Wu YH, Aksan N, Rizzo M, et al. Measuring listening effort: driving simulator versus simple dual-task paradigm. *Ear Hear*. 2014; 35:623–632. [PubMed: 25083599]
- Zekveld AA, Kramer SE. Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*. 2014; 51:277–284. [PubMed: 24506437]
- Zekveld AA, Kramer SE, Festen JM. Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear*. 2010; 31:480–490. [PubMed: 20588118]

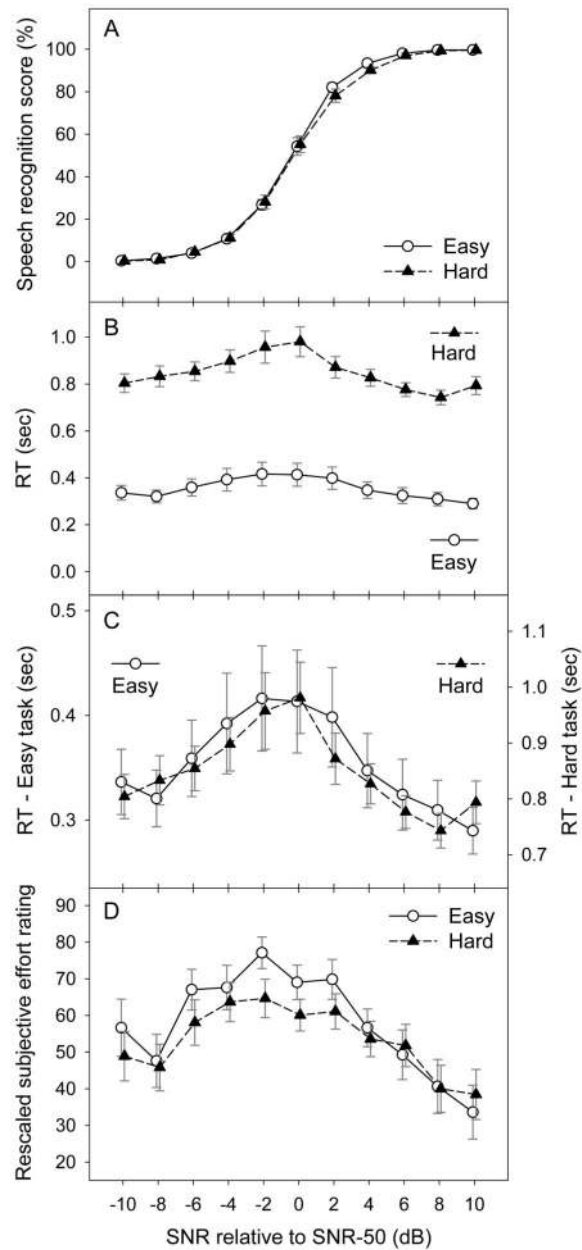


Figure 1.

Speech recognition score (1A), reaction time (RT) of the secondary task (1B and 1C), and subjective listening effort rating (D) averaged across participants as a function of signal-to-noise ratio (SNR) of Experiment 1. In Figure 1C the y-axes of the RT curves are rescaled so that the curves of the easy task (refer to the left y-axis) and the hard task (refer to the right y-axis) have similar peak heights in the figure. Error bars = 1 SE.

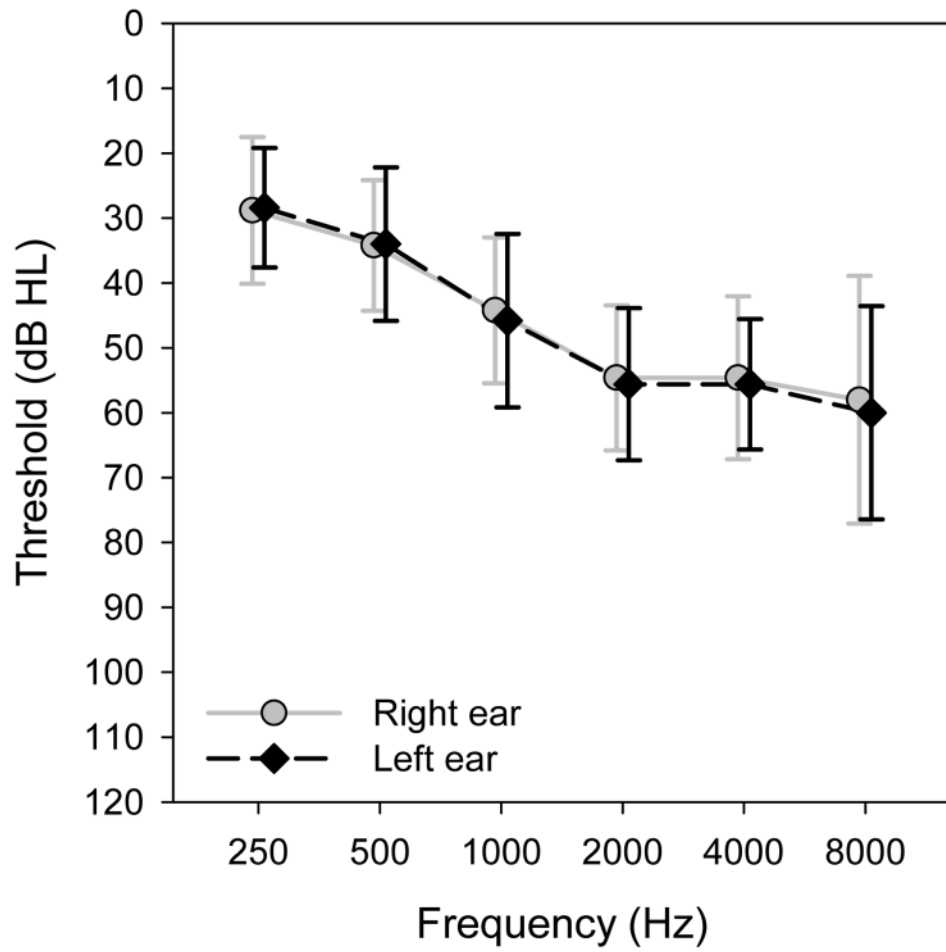


Figure 2. Average audiograms of study participants in Experiment 2. Error bars = 1 SD.

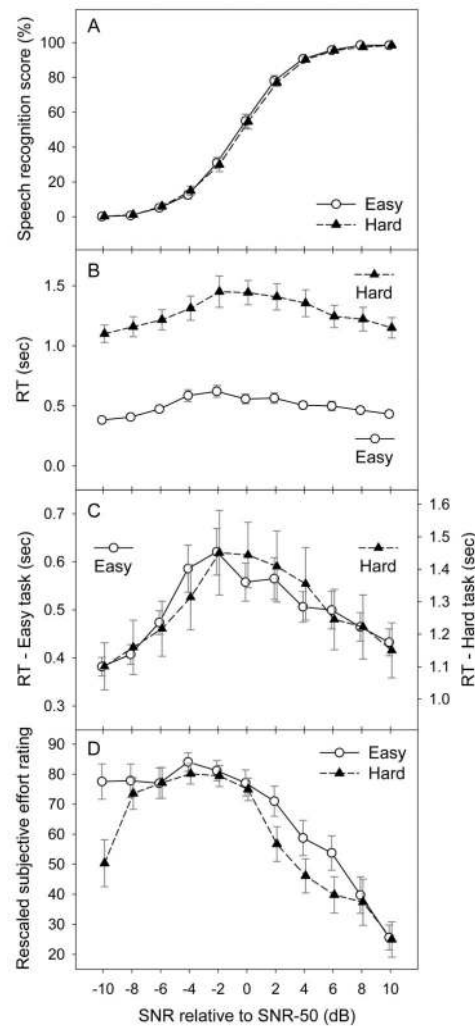


Figure 3. Speech recognition score (3A), reaction time (RT) of the secondary task (3B and 3C), and subjective listening effort rating (3D) averaged across participants as a function of signal-to-noise ratio (SNR) of Experiment 2. In Figure 3C the y-axes of the RT curves are rescaled so that the curves of the easy task (refer to the left y-axis) and the hard task (refer to the right y-axis) have similar peak heights in the figure. Error bars = 1 SE.

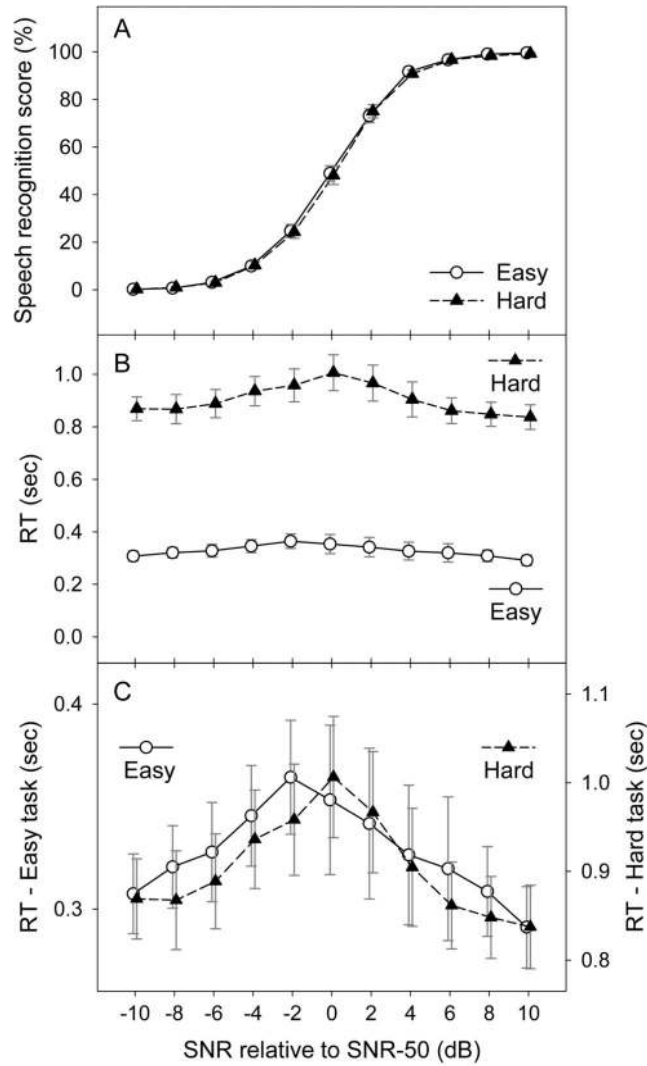


Figure 4.

Speech recognition score (4A) and reaction time (RT) of the secondary task (4B and 4C) averaged across participants as a function of signal-to-noise ratio (SNR) of Experiment 3. In Figure 4C the y-axes of the RT curves are rescaled so that the curves of the easy task (refer to the left y-axis) and the hard task (refer to the right y-axis) have similar peak heights in the figure. Error bars = 1 SE.

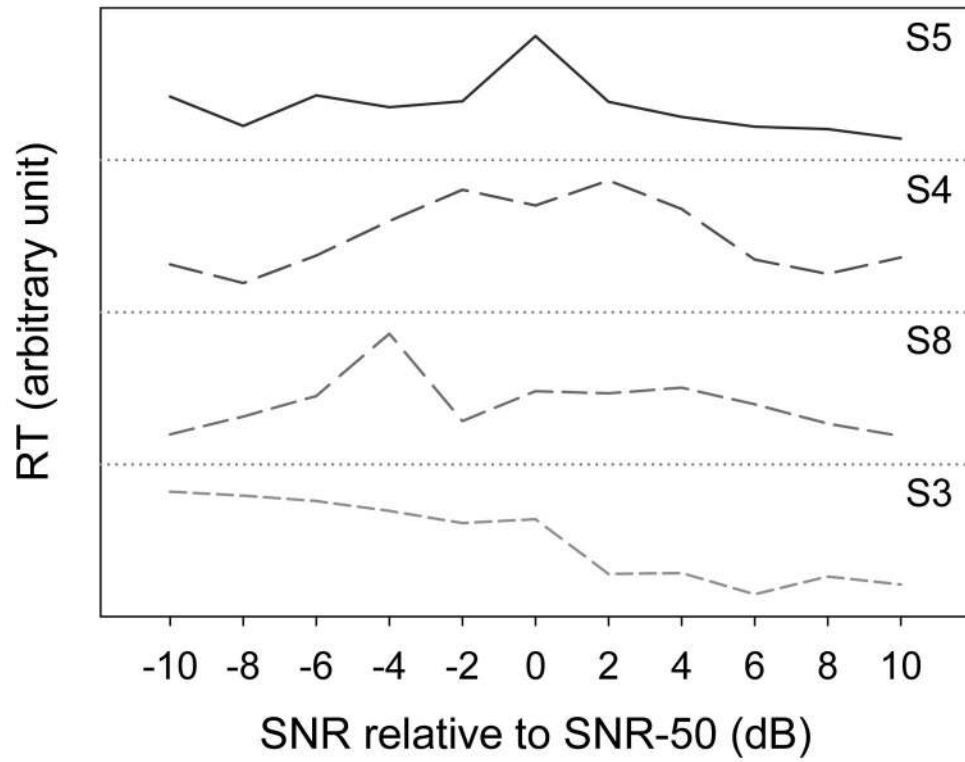


Figure 5. Reaction time (RT) of the hard secondary task as a function of signal-to-noise ratio (SNR) for four participants (S5, S4, S8, and S3) in Experiment 3. The curves are rescaled to have similar peak heights in the figure.

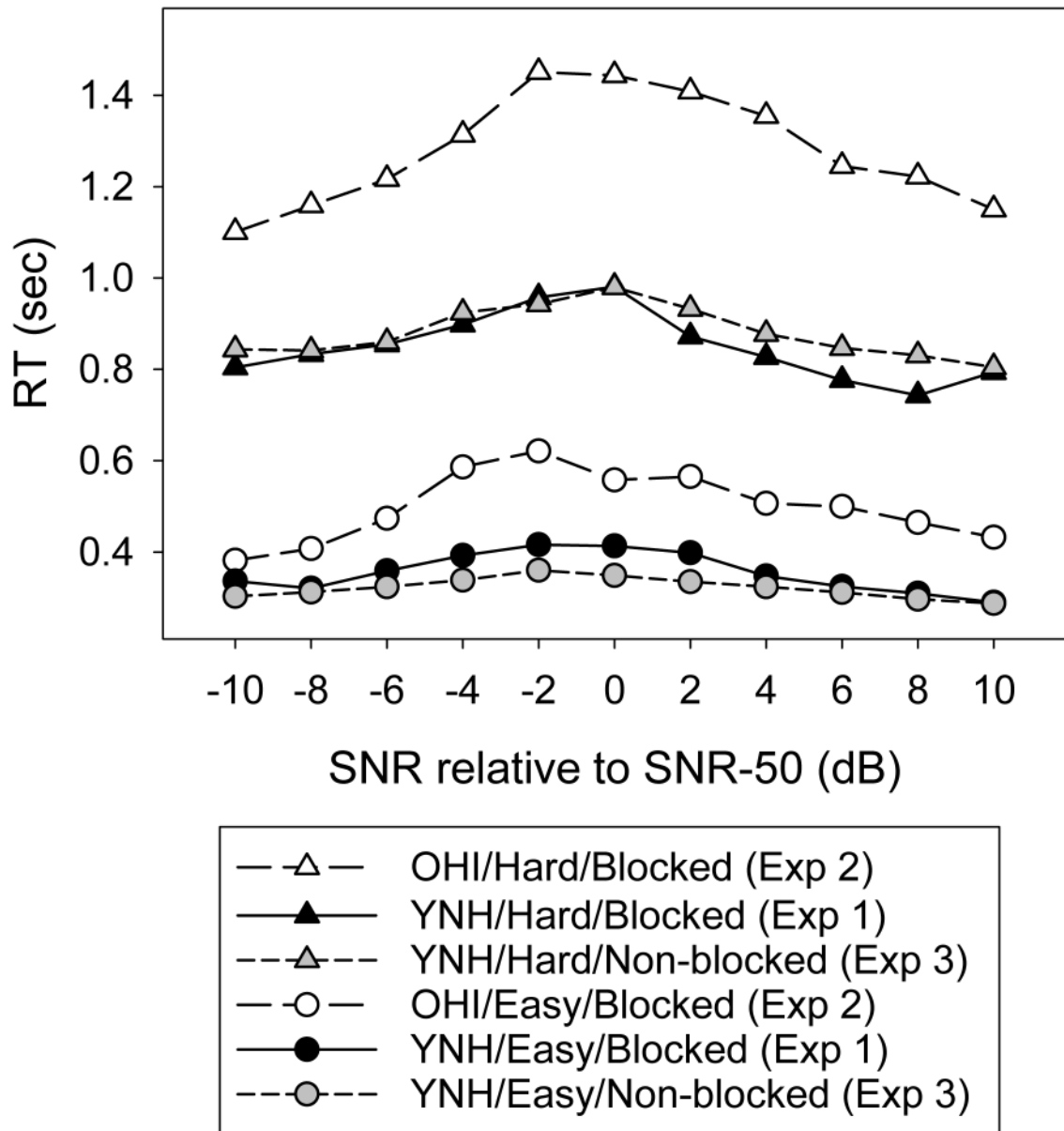


Figure 6. Reaction time (RT) of the secondary task averaged across participants as a function of signal-to-noise ratio (SNR) of Experiments 1, 2, and 3. YNH: younger adults with normal hearing; OHI: older adults with hearing impairment; Easy/Hard: the type of the secondary task; blocked/non-blocked: SNR presentation order; Exp: experiment.