Biomedical Engineering Letters
https://doi.org/10.1007/s13534-023-00315-6

ORIGINAL ARTICLE

Pupillometry as a window to detect cognitive aging in the brain

Jiae Kim[1] · **Jiyeon Lee**[2] · **Sang Beom Jun**[2,3,4] · **Jee Eun Sung**[1]

Received: 2 June 2023 / Revised: 24 July 2023 / Accepted: 22 August 2023

Abstract

This study investigated whether there are aging-related differences in pupil dilation (pupillometry) while the cognitive load is manipulated using digit- and word-span tasks. A group of 17 younger and 15 cognitively healthy older adults performed digit- and word-span tasks. Each task comprised three levels of cognitive loads with 10 trials for each level. For each task, the recall accuracy and the slope of pupil dilation were calculated and analyzed. The raw signal of measured pupil size was low-pass filtered and interpolated to eliminate blinking artifacts and spike noises. Two-way ANOVA was used for statistical analyses. For the recall accuracy, the significant group differences emerged as the span increases in digit-span (5- vs. 7-digit) and word-span (4- vs. 5-word) tasks, while the group differences were not significant on 3-digit- and 3-word-span tasks with lower cognitive load. In digit-span tasks, there was no aging-related difference in the slope of pupil dilation. However, in word-span tasks, the slope of pupil dilation differed significantly between two groups as cognitive load increased, indicating that older adults presented a higher pupil dilation slope than younger adults especially under the conditions with higher cognitive load. The current study found significant aging effects in the pupil dilations under the more cognitive demanding span tasks when the types of span tasks varied (e.g., digit vs. word). The manipulations successfully elicited differential aging effects, given that the aging effects became most salient under word-span tasks with greater cognitive load especially under the maximum length.

Keywords Aging · Cognitive load · Short-term memory tasks · Cognition measuring tool · Pupillometry

1 Introduction

Research has demonstrated that mental activities can be captured using physiological tools measured at the brain level (e.g., electroencephalogram, EEG) or at the peripheral level (e.g., pupillometry). More recently, pupillometry has gained considerable attention as a measure of higher-level cognitive-linguistic processing, demonstrating that the pupil dilation from pupillometry can be a biological window into a higher level of mental activities and effort. In general, as the difficulty of cognitive tasks increases, the pupils tend to dilate [1]. Conversely, when the cognitive load of a task stabilizes, the pupils tend to decrease or contract in a rest [2]. Moreover, when the processing demands exceed ones' cognitive capacity, the pupil size decreases [3]. Thus, cognitive load can influence the variability in the pupil dilation and contraction responses between individuals. Furthermore, due to the practical and clinical utility of pupillometry as a non-invasive method, it has been applied to diverse populations of interest such as infants [4] and clinical populations with neurological disorders [5] along with various tasks measuring different domains of mental activities such as memory [6, 7], language [8, 9], and emotion [10]. The current study employed the pupillometry to examine real-time evidence of aging-related changes to short-term memory tasks as cognitive load increased.

Jiae Kim and Jiyeon Lee contributed equally to this work and share first authorship.

✉ Sang Beom Jun
juns@ewha.ac.kr

✉ Jee Eun Sung
jeesung@ewha.ac.kr

[1] Department of Communication Disorders, Ewha Womans University, Seoul, Republic of Korea

[2] Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul, Republic of Korea

[3] Graduate Program in Smart Factory, Ewha Womans University, Seoul, Republic of Korea

[4] Department of Brain and Cognitive Sciences, Ewha Womans University, Seoul, Republic of Korea

Short-term memory (STM) refers to the short-term storage buffer that maintains information units [11, 12], and it has been traditionally measured using digit-span tasks [13–15], in which participants are required to recall the digits that they heard. Considering that STM tasks are regarded as an index of cognitive capacity, aging-related cognitive decline is often measured using STM tasks. It is well-known that older adults do not perform as well in the STM tasks as younger adults, suggesting that cognitive capacity reduces with age [16, 17]. However, it is still controversial whether the aging-related reduction in cognitive capacity affects real-time interpretive information processing or post-interpretive levels of processing [18; series of the debates]. As a real-time measure of cognitive load using STM tasks, the pupillometry has been used in several studies with diverse population groups such as normal populations from children to aging or neurogenic populations. Johnson et al. [19] applied the pupillometry using digit-span tasks to examine real-time differences in STM capacity between children (mean age = 10.6 years) and adults (mean age = 27.5 years). As a result, the pupils of children dilated similarly to those of adults under the 6-digit condition. However, the group differences emerged as the digit-span increased up to 9-digit, demonstrating that pupil dilation continuously increases in adults as the digit load grows, whereas children's pupil contracted when the span exceeded 6-digit. The authors suggested that the pupillometry can be used as a measure to reveal developmental changes in cognitive processing.

Another line of research regarding STM and pupillometry derives from evidence collected on aging populations. Piquado et al. [20] investigated aging-related changes in pupil dilation as a function of cognitive load, as measured by digit-span recall and tasks. Pupil dilation gradually increases as the digit-span increases in both younger and older adults. However, no significant aging-related differences were observed in pupil dilation while the participants listened to the digits. In contrast, significant group differences emerged during the retention period, which refers to the time interval after the subjects finished listening and before they were required to recall. The results suggested that the impacts related to aging depend on the time point when the pupil dilation is measured, indicating that aging-related differences are predominant at the stage of actively retaining the information rather than passively listening to the digits.

More recently, pupillometry was applied to at-risk aging populations like those with mild cognitive impairment (MCI), a transitional stage between normal aging and dementia [21]. Amnestic MCI has been reported as carrying a higher rate of conversion to Alzheimer's disease than other types [22]. Granholm et al. [21] investigated the pupillary responses of participants listening to digit-span tasks by comparing cognitively normal to amnestic and non-amnestic MCI groups. Amnestic MCI individuals demonstrated greater pupil dilation than non-amnestic MCI and the cognitively normal control group. Authors argued that pupil responses are sensitive enough to classify the amnestic MCI group as high risk, compared to normal aging and non-amnestic MCI groups. Thus pupillometry using STM tasks may serve as a biomarker for the early detection of neurodegenerative disease.

According to previous studies, pupillometry offers the potential for early detection of aging-related cognitive decline. However, it should be noted that Piquado et al. [20] found the effects related to aging only in the retention period, not in the listening stage, while the amnestic MCI group showed differences in pupil dilation while listening to the digits [21]. Those differences seem to be related to individual differences in cognitive capacity, meaning that real-time differences in pupil dilation could become more distinct as cognitive capacity reduces. We speculate that the non-significant results in the listening stage from Piquado et al. [20] were possibly the result of digit-span measures that were not cognitively taxing enough to induce aging-related differences. Therefore, to increase cognitive demands, we administered word-span tasks added to the measurement with digit-span tasks to examine whether aging-related cognitive decline can be observed using pupillometry. The word-span systematically increases, and the participant must recall the words that are presented in a given span. It is identical to the way in which the digit-span is administered [23]. The differences arise in the unit that they recall: digits vs. words. Word-span tasks required the activation of the semantic features of language domains, and the demands on linguistic memory are known to increase the cognitive burdens of the tasks [24]. Researchers have found that people demonstrated greater difficulties with word-span rather than digit-span tasks [25].

The current study investigates whether the aging-related differences emerged in pupil dilation as the cognitive load was systematically manipulated using digit and word-span tasks. We hypothesized that the word-span may induce significant aging effects during the span tasks because it requires greater cognitive and linguistic demands than the digit-span. Specific aims of the study include (1) to examine age group differences in the recall accuracy of digit- and word-span tasks as a function of cognitive load, and (2) to explore aging-related differences in the slope of pupil dilation as the cognitive load systematically increases.

2 Materials and methods

2.1 Participants

A total of 32 individuals participated in the study with two age groups: younger adults (n = 17; mean age = 24.35 ± 3.59 years) and older adults (n = 15; mean age = 63 ± 2.88 years). All of them met the following inclusion and exclusion criteria: (1) native Korean speakers, (2) no history of mental or neurological disorders, as per the screening exclusion criteria [26], (3) no problems with vision or hearing stimuli above 30 dB [27], which was required to perform the tasks, (4) no problems with listening and comprehending language, and (5) more than nine years of education. The mean years of education did not differ significantly between the groups (F(1.31) = 0.023, p > .05). Both groups should also meet the inclusion criteria in mental state, verbal learning, and depression scale (**Supplementary Table 1**). The studies involving human participants were reviewed and approved by the Institutional Review Board of Ewha Womans University (EWIRB-202205-0032-01). The participants provided their written informed consent to participate in this study.

2.2 Cognitive tasks

The cognitive tasks include digit-span tasks and word-span tasks that are simple behavioral measures of verbal STM capacity. In both tasks, the subjects hear increasingly longer sequences of digits or words and immediately attempt to recall them in the same order. Digit-span tasks referred to Digit Forward Recall of K-WAIS (Korean Wechsler Adult Intelligence) [28], and word-span tasks referred to the Korean Word Forward Test demonstrated in the study by Sung [29]. All word lists were prepared, based on the vocabulary frequency of Yonsei Corpus I [30] according to the criteria presented by Sung [29]. Table 1 shows example of two tasks with different spans.

Digit-span tasks consisted of a total of 30 lists. These lists were divided into spans of 3-, 5-, and 7-digit, each consisting of 10 lists. The 30 lists were further organized into two blocks, with each block consisting of 15 lists. Five lists were randomly selected for each span to make one block, and the other block was made using the remaining lists in each span. The sequence of span tasks was pseudo-randomized to ensure that the same span items weren't presented three times consecutively with the same span. Additionally, word-span tasks consisted of a total of 30 lists. These lists were divided into spans of 3-, 4-, and 5-word, each consisting of 10 lists. The blocks for word-span tasks were created in the same manner as those for digit-span tasks. The lists within a block were created to have different initial phonemes, and any consecutive occurrence of the same word three or more times was controlled. Participants were given the lists in the same order, in which the order of items was randomized for each list.

Table 1 Examples of digit-span tasks and word-span tasks (The word span is originally in Korean, but it written in English to aid understanding.)

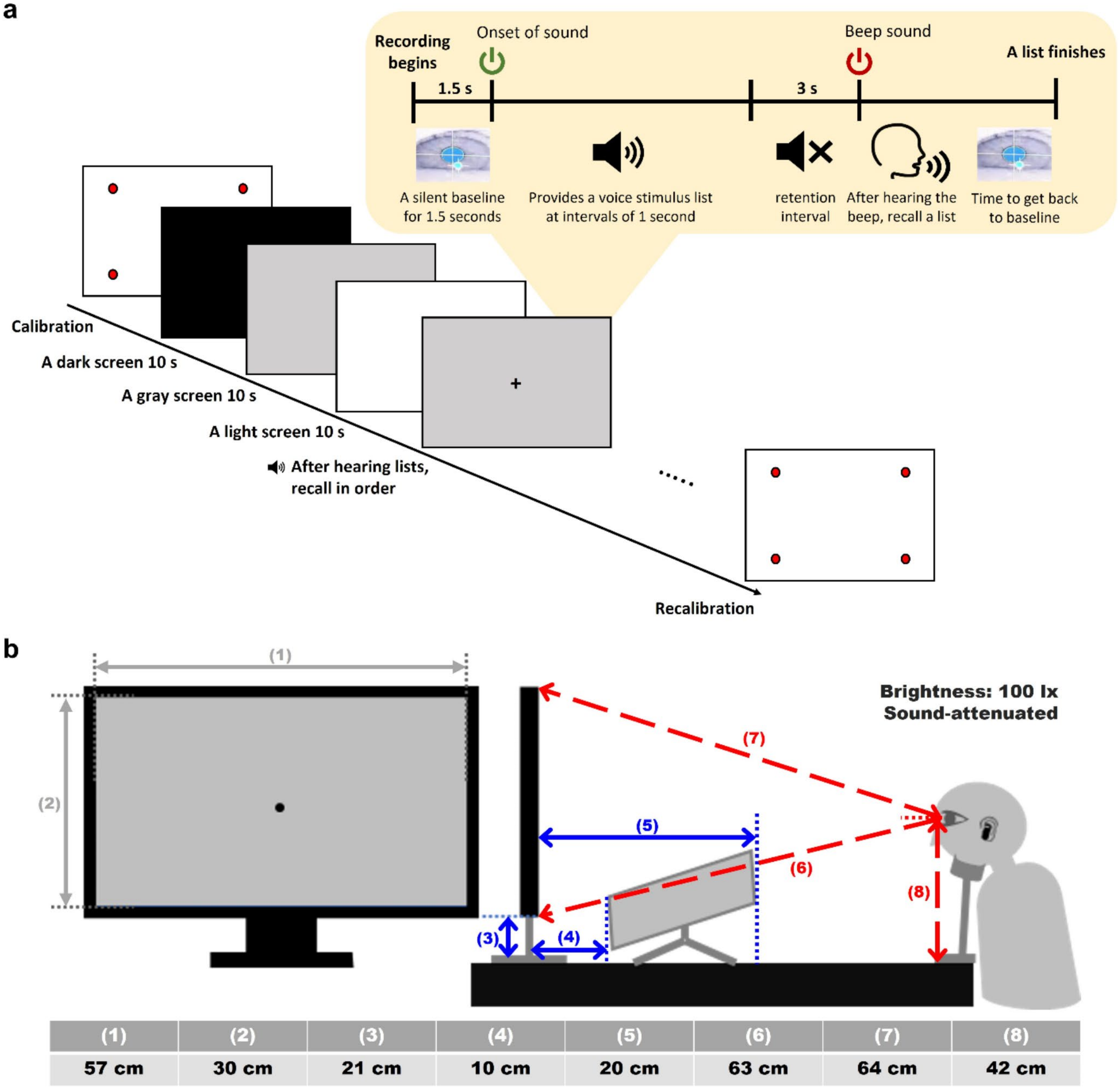
span	Examples of digit-span tasks	span	Examples of word-span tasks
3-digit	2–6 – 9	3-word	apple – mirror – arm
5-digit	6–8 – 3–7 – 1	4-word	school – money – farmer – rat
7-digit	1– 6–4 – 8–3 – 5–9	5-word	eye – wing – nose – tower – pants

2.3 Experimental procedure

To calibrate the individual variability of the pupil size, every participant's pupil size was measured and converted into the pixel size. After calibration, the pupil sizes of each participant were measured every 10 s in the order of dark, gray, and light to measure physiologically maximum (dark phase) and minimum (light phase) pupil size. Next, the participants listened to cognitive tasks at the same volume using earphones while looking at the fixation cross (+) on the screen. At the end of each list, a beep sound was provided after three seconds, and they recalled in the same order without time limit. Every fifteen stimulus lists, participants received a three- to five-minute break to reduce eye fatigue (Fig. 1a).

Prior to the experiment, all participants were confirmed to have no issues in perceiving stimuli at a level of 30 dB and higher. All stimuli were recorded by professional Korean voice actresses and standardized to a sound pressure level of 70–75 dB using GoldWave (GoldWave Digital Audio Editor, version 6.57). To return the dilated pupil to its innate size, each list began with a 1.5-second silent baseline. Also, at the end of each list, a beeping sound was delivered as a recall start signal following a three-second retention interval. During the voice stimuli for tasks, the interval was one second between each digit or word. Digit-span tasks took 8.0, 10.8, and 13.6 s for 3-, 5-, and 7-digit lists, and the word-span task took 8.8, 10.2, and 11.6 s for 3-, 4-, and 5-word lists, respectively.

Since pupillometry is very sensitive to environmental factors, including light, emotional and physical stimuli, and distance between the pupillometry device and the participant, the whole tasks were performed in a room with specific conditions and participants were required to maintain specific stances (Fig. 1b). Pupil sizes and eye movements were recorded binocularly using an infra-red video-oculography system (Eyelink portable Duo, SR Research), and the entire experiment was performed with Experimental Builder software (SR Research, Canada). The pupil sizes were sampled



(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
57 cm	30 cm	21 cm	10 cm	20 cm	63 cm	64 cm	42 cm

Fig. 1 Display of an overall experimental paradigm. **(a)** Experimental procedures. **(b)** Environmental conditions. The participant was required to continuously stare at a black cross ($2\times2\ \text{cm}^2$) on a gray screen (3 fL) during experiments for visual fixing. An environment with consistent ambient light (100 lx, whole room), a sound-attenuated room, and a fixed distance between participants and the pupillometry device were maintained for the experiment

at 1 kHz, and the positions were calibrated based on nine points at the beginning of the experiments [31].

2.4 Data analysis

In this study, two different types of data were obtained to detect the aging effect: recall accuracy as an offline measure and pupil dilation as an online measure.

2.4.1 Recall accuracy

The recall accuracy was calculated from each list of the tasks. Recall accuracy serves as an offline measure, involving the recording of participants' behavioral and physiological data during the experiment for subsequent analysis. It was determined as the percentage of lists that were correctly recalled in both items and order for each cognitive task (digits or words).

$$Recall\,accuracy\ [\%] = \frac{Number\,of\,correct\,lists}{Total\,number\,of\,lists} \times 100$$

2.4.2 Pupil size

To eliminate innate differences in pupil size between participants and trials, absolute values of pupil size were normalized with respect to the maximum pupil size, which are referred to as pupil dilation (PD) in this paper. Maximum pupil size was obtained when gazing at a dark screen. The data were segmented and averaged by the duration of each task. Thus the normalized value, PD, was utilized to represent pupil size.

$$Pupil\,dilation\,(PD)\ [\%] = \frac{Measured\,pupil\,size\ [mm]}{Maximum\,pupil\,size\ [mm]} \times 100$$

The slope of pupil dilation (PD) was also calculated from each task list, serving as an online measure that involves collecting real-time data. The slope of PD was determined by calculating the difference in pupil size per second between the first and last span, which indicates the change in pupil dilation during that specific time period, in order to compare different span tasks in the same time scale. Instead of comparison with direct PD, the slope of PD was used for statistical analyses because this parameter from linear regression could represent more effective for complex psycholinguistic research [32].

$$Slope\,of\,PD\ [\%\,/\,\mathrm{sec}] = \frac{(Average\,of\,PD\,of\,last\,span) - (Average\,of\,PD\,of\,first\,span)\ [\%]}{Total\,duration\,of\,a\,list\ [\mathrm{sec}]} \times 100$$

To ensure the acquisition of reliable pupil data, a series of signal processes were employed as shown in Fig. 2. The analyses were conducted using custom-made MATLAB scripts (Matlab 2020b, Mathworks, Natick, MA).

After the recording of pupil size, data processing was performed to remove noise and transform it into a suitable format for analysis (Fig. 2). The pupil dilation data was analyzed on a per-list basis. There were two types of noise in the pupil data: blink-related noise and saccade-related noise. To address blink-related noise, such as blink artifacts and spikes, a method leveraging the specific characteristics of each type of noise was used for identification and removal. The gaps created by the removal of blink artifacts and spikes were filled using linear interpolation. For saccade-related noise, 10 Hz fourth-order Butterworth filter with a zero-phase shift lowpass filter and local regression (LOESS; locally estimated scatterplot smoothing) were applied to effectively remove the noise [33]. Following the noise reduction process, the pupil diameter and the slope of the pupil diameter were calculated, and statistical analysis was conducted on the data.

2.5 Statistical analyses

A two-way mixed repeated-measures analysis of variance (ANOVA) was performed to determine if there were any significant differences in pupil responses between younger and older groups while performing the cognitive tasks. Statistical significance was indicated by p-value under 0.05. Mauchly's sphericity test was performed for the repeated measures ANOVA, and when the sphericity assumption was not satisfied, the degrees of freedom and F value corrected by Greenhouse-Geisser were analyzed. The data was analyzed with SPSS ver. 26 (SPSS Inc., Chicago, IL, USA).

3 Results

3.1 The recall accuracy

3.1.1 Digit-span tasks

In digit-span tasks, younger adults correctly recalled 100%, 98.23%, and 86.47% and older adults correctly recalled 99.33%, 90.67%, and 65.33% of the 3-, 5-, and 7-digit conditions, respectively (Fig. 3a). As a result of 2 (group) × 3 (length of span) ANOVA, the main effect of the groups was significant ($F(1,30) = 14.949$, $p = .001$). That is, the recall accuracy of the older group ($M = 85.111$, $SE = 1.846$) was significantly lower than that of the younger group ($M = 94.902$, $SE = 1.734$). The main effect for the length of span (3-, 5-, 7-digit) was significant ($F(1.147,34.421) = 33.165$, $p = .000$). As a result of post-hoc testing using Bonferroni, the differences between 3-digit and 5-digit, 3-digit and 7-digit, 5-digit and 7-digit conditions were all significant ($p < .01$). There was a significant two-way interaction between the group and length of span ($F(1.147,34.421) = 5.765$, $p = .018$). For post-hoc analyses of the significant interaction, one-way ANOVAs between younger and older groups were performed for each length of span. The results showed that the difference between the two groups was not significant in 3-digit conditions with low cognitive load ($F(1,31) = 1.138$, $p = .295$). However, as the cognitive load increased, such as with 5-digit ($F(1,31) = 10.217$, $p = .003$) and 7-digit ($F(1,31) = 8.783$, $p = .006$) conditions, the recall accuracy of the older group significantly decreased compared to the younger group.

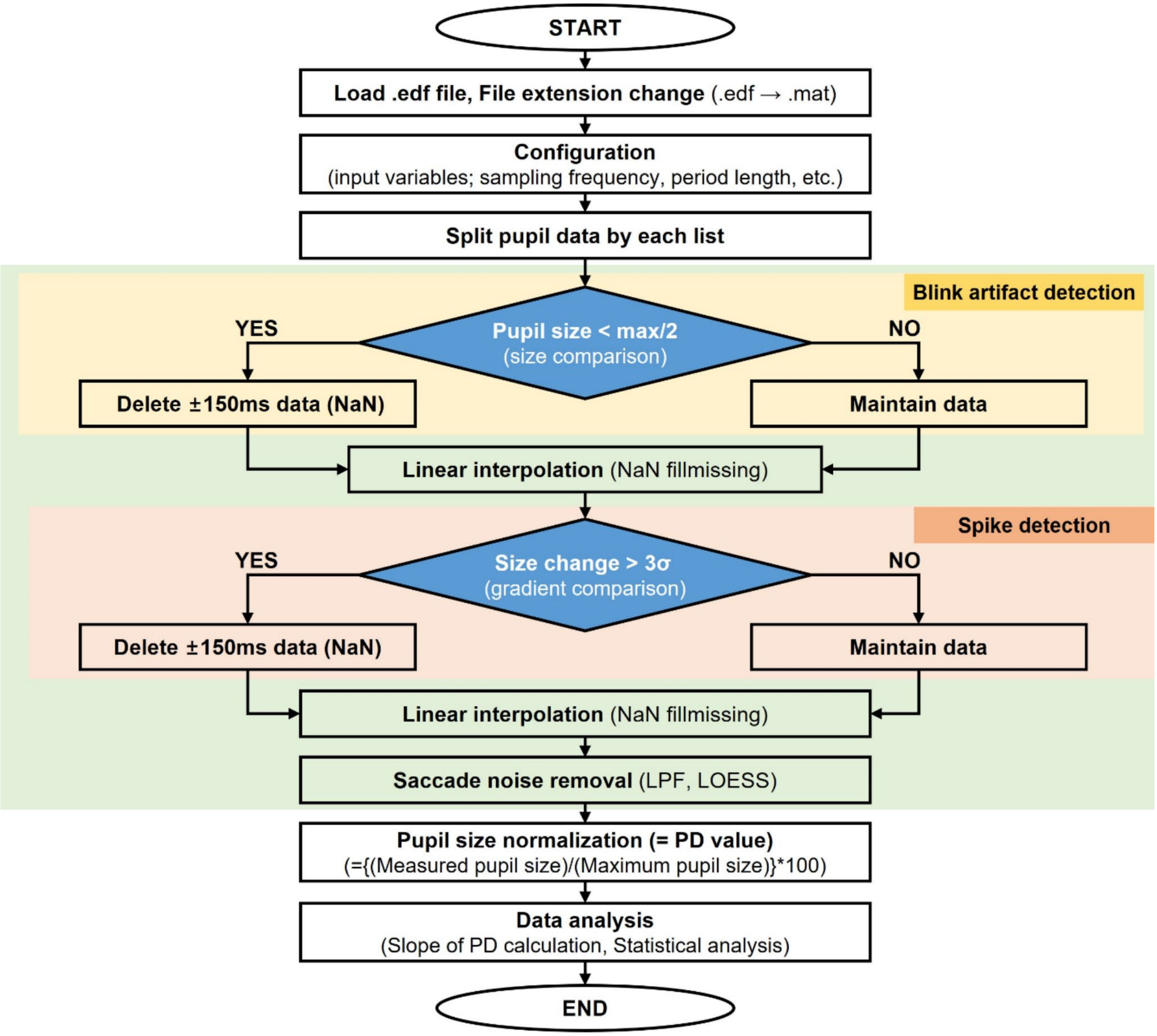


Fig. 2 Flowchart of data analysis. σ = standard deviation of pupil size. The blinking of the eye can generate two types of noise: blink artifacts, which detect excessively low pupil size values (depicted by the yellow box), and spikes, which detect rapidly changing pupil size caused by variations in light intensity (depicted by the orange box). The green box represents the process of noise removal

3.1.2 Word-span tasks

In word-span tasks, younger adults correctly recalled 100%, 98.23%, and 87.64% and older adults correctly recalled 98%, 78%, and 32% of the 3-, 4-, and 5-word conditions, respectively (Fig. 3b). As a result of 2 (group) × 3 (length of span) ANOVA, the main effect of the groups was significant (F(1,30) = 86.049, p = .000). The recall accuracy of the older group (M = 69.33, SE = 2.04) was significantly lower than that of the younger group (M = 95.29, SE = 1.91). The main effect for the length of span (3-, 4-, 5-word) was significant (F(2,60) = 107.79, p = .000). As a result of post-hoc testing using Bonferroni, the differences between 3-word and 4-word, 3-word and 5-word, 4-word and 5-word conditions were all significant (p = .000). There was a considerable two-way interaction between group and length of span (F(2,60) = 49.033, p = .000). For post-hoc analyses of the significant interaction, one-way ANOVAs between younger and older groups were performed for each length of span. The results showed that the difference between the two groups was not significant in 3-word conditions with low cognitive load (F(1,31) = 2.173, p = .151). However, as the cognitive load increased, such as with 4-word (F(1,31) = 26.549, p = .000) and 5-word (F(1,31) = 84.643, p = .000) conditions, the recall accuracy of the older group significantly decreased compared to the younger group.

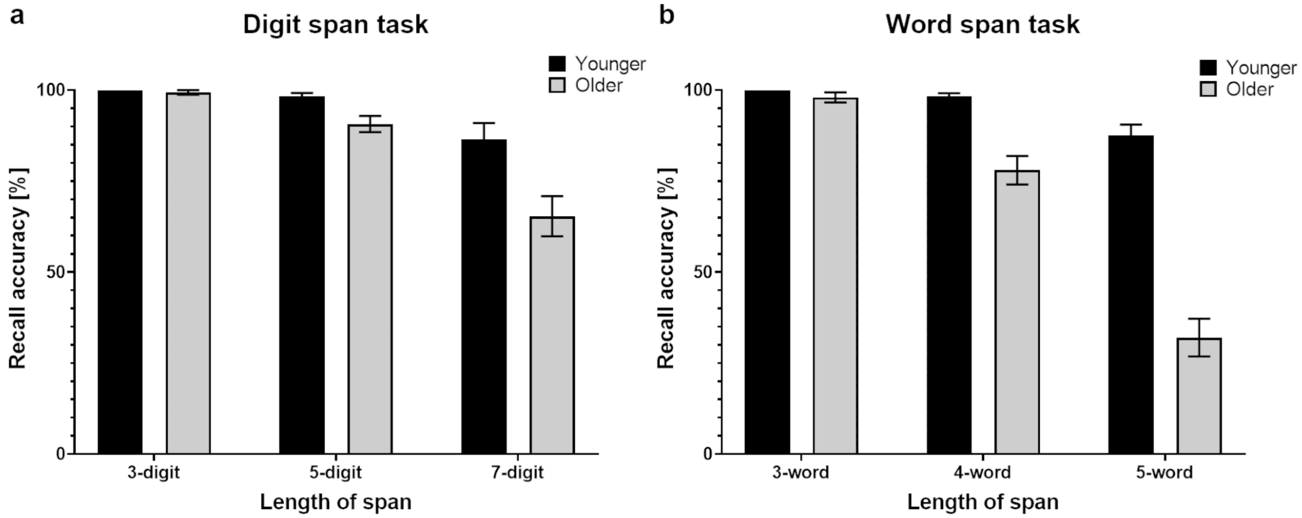


Fig. 3 The recall accuracy of **(a)** Digit-span tasks **(b)** Word-span tasks. Results for younger adults were shown in black, and for older adults were shown in gray. Each error bar indicates standard error

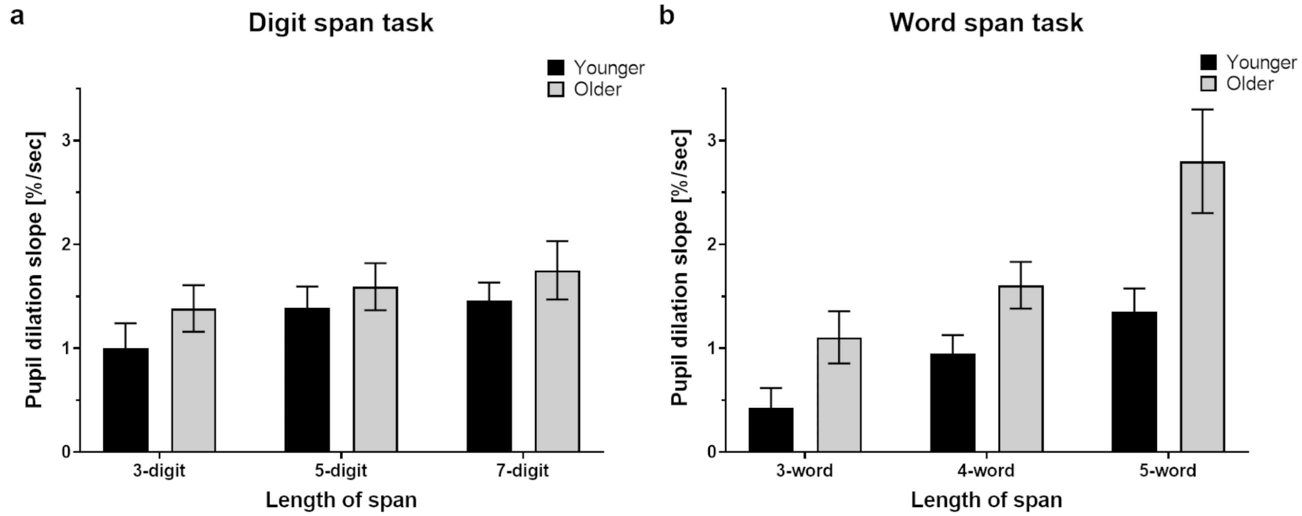


Fig. 4 The pupil dilation slope of **(a)** Digit-span tasks **(b)** Word-span tasks. Results for younger adults were shown in black, and for older adults were shown in gray. Each error bar indicates standard error

3.2 The slope of pupil dilation

3.2.1 Digit-span tasks

The pupil dilation slopes of the younger adults were 1.00%/sec, 1.39%/sec, and 1.46%/sec and those of the older adults were 1.38%/sec, 1.59%/sec, and 1.75%/sec in 3-, 5-, and 7-digit conditions of digit-span tasks, respectively (Fig. 4a). As a result of 2 (group) × 3 (length of span) ANOVA, the main effect for the length of span (3-, 5-, and 7-digit) was significant (F(2,60) = 8.162, p = .001). Post-hoc comparisons using Bonferroni revealed that there were significant differences between 3- and 5-digit (p = .037) and between 3- and 7-digit (p = .005), whereas differences between 5- and 7-digit conditions were not significant, p = 536. However, there were not significant effects for the group (F(1,30) = 0.906, p = .349) and the two-way interaction (F(2, 60) = 0.361, p = .699).

3.2.2 Word-span tasks

The pupil dilation slopes of the younger adults were 0.43%/sec, 0.95%/sec, and 1.35%/sec and those of the older adults were 1.10%/sec, 1.61%/sec, and 2.80%/sec in 3-, 4-, and 5-word conditions of word-span tasks, respectively (Fig. 4b). As a result of 2 (group) × 3 (length of span) ANOVA, the main effect among the three span conditions (3-, 4-, and 5-word) was significant (F(2,60) = 27.514, p = .000). As a result of post-hoc testing using Bonferroni, the differences between the 3-word and 4-word (p = .004), 3-word and 5-word (p = .000), 4-word and 5-word (p = .000) conditions were all significant. Also, the main effect of the group was significant (F(1,30) = 7.262, p = .011). The two-way interaction between the span and group was statistically significant (F(2,60) = 3.189, p = .048). The interaction contrasts were conducted as post-hoc analyses using the LMATRIX and MMATRIX commands. As a result, the group differences between the 4-word and 5-word conditions were significant (p = .035). However, the group differences were not significant for both comparisons between 3-word and 5-word (p = .07) and between 3-word and 4-word (p = .945) conditions. The significant interaction is likely due to the fact that the slope of pupil dilation differentially increased for the older group when the word-span had the highest demands (e.g., 5-word) compared to the younger adults.

4 Discussion

The current study investigated aging-related differences in pupil responses by systematically manipulating the cognitive load of two STM tasks. Regarding the recall accuracy, older adults demonstrated significantly lower performance than the younger adults across the tasks. These results were consistent with previous studies showing that cognitive capacity, as indexed by working memory or STM, declines with aging [16, 17, 34, 35]. The recall accuracy decreased as the length of span increased across the groups, consistent with previous findings that suggested that performance degraded as task demands increased [20, 36]. The two-way interactions between the age groups and length of span were significant for both digit- and word-span tasks. Post-hoc analyses revealed that there were no significant differences between the age groups in the shortest span (3-digit and 3-word), whereas group differences emerged as the span increased. The results indicate that cognitive capacity was taxed as the span increased, and these effects had specifically greater impact on the older group. The results were consistent with previous findings suggesting that cognitive capacity reduces as people age. In addition, in these studies, the limited capacity effects emerged under the cognitively challenging conditions [37–39].

The slopes of pupil dilation, derived from real-time measurement, showed significant aging-related effects only during word-span tasks, with older adults having significantly higher values than younger adults. The current results are consistent with previous studies that showed no age group differences in the slope of pupil dilation while listening to stimuli lists in digit-span tasks [20], indicating that digit-span tasks might not be a sensitive enough measure to reveal the aging-related effects in the slope of pupil dilation. However, there is some counterevidence that suggests that digit-span tasks were sufficient enough to produce group differences in pupil dilation between children and younger adults [19], and between the MCI and normal-aging groups [21]. We speculate that the discrepancy between the current findings and those of previous studies is likely driven from the characteristics of the groups that are compared using pupillometry for digit-span tasks. Previous studies found significant group differences by comparing groups of vastly different ages (e.g., children to young adults) or clinical stages (e.g., normal aging vs. MCI as a high-risk group for dementia). However, studies that included normal younger and older adults did not observe significant group differences in pupil dilation during digit-span tasks [20]. The results suggest that group differences in cognitive capacity may account for the inconsistent findings across the studies. In other words, group differences in pupil dilation seem to become salient when the groups with clearly reduced capacity are compared.

Unlike digit-span tasks, we found significant aging-related effects in the slope of pupil dilation under word-span tasks. Previous research reported that word-span tasks are more difficult than digit-span tasks, given that word-span tasks involve greater cognitive loads on the lexical-semantic activation compared to the digit tasks [24, 25, 40, 41]. In other words, digits represent abstract concepts, which reduces the need for semantic processing demands. However, the word-span task requires complex cognitive processing, such as understanding and processing the meaning, analogy, and association of words. Words can also vary in length, number of syllables, and distance between syllables [41]. This variability can cause confusion in the process of remembering and retrieving stimuli, increasing the difficulty of the task. Also, the older group may exhibit lower performance in word-span tasks than the younger group due to slower information processing speed, limited working memory capacity, reduced cognitive control, difficulties in word comprehension, and challenges in processing meaning [40]. Kahneman and Beatty [2] examined pupil dilation in younger adults using both digit- and word-span tasks. They investigated pupil dilation during the tasks, in which both two STM tasks increased from a span of 3 up to a span of 7. The results suggested that participants experienced greater pupil dilation under the word-span condition, rather than the digit-span condition, which is consistent with the current results. Overall, these findings suggest that word-span tasks contain more cognitive load than digit-span tasks, and these increased demands are reflected in the real-time measures of pupil dilations.

Consistent with the findings that suggest that the age group effects emerged only for word-span tasks and interaction between the group and the length of word-span. Group differences between younger and older adults became significant between 4- and 5-word length of span, indicating that older adults experienced differentially greater pupil dilations compared to the younger group as the length of word-span increased. The results are consistent with the hypothesis that the increased length under the more cognitively demanding tasks (e.g., word-span) taxed the group with reduced cognitive capacity (e.g., the older group). It is interesting to note that significant interaction was only under word-span tasks, although digit-span (up to 7) has a greater length than the word-span (maximum span = 5). Despite the shorter length of the span for the word-span, the tasks resulted in significant aging-related effects as the span increased. This suggest that the cognitive load is sufficient for the word-span due to the greater lexical-semantic load imposed on the word by compensating for the length of span. These findings imply that the cognitive load needs to be considered from the perspectives of both the length and computational loads imposed on the tasks, as consistent with the working memory capacity theory by Just and Carpenter [42].

In conclusion, the current study found significant aging effects in the pupil dilations under the more cognitive demanding span tasks when task demands were manipulated by varying the types of span tasks (e.g., digit vs. word). The manipulations successfully elicited differential aging effects depending on the span task types, given that the aging effects became most salient under word-span tasks with greater cognitive load especially under the maximum length. The results are consistent with the theoretical underpinnings, which posit that cognitive capacity is limited as people age, and the effects of limited capacity emerged when their resources are taxed enough to be exceeded [42]. Furthermore, the current study confirmed that these effects manifest themselves in real-time processing, reflected in the changes to pupil dilation. More studies are required to manipulate a diverse range of cognitive loads imposed by different task types to study the effects of aging using pupillometry in real-time processing. These findings may be further applied to aging populations at risk for dementia (e.g., MCI), meaning that pupillometry could be used as a diagnostic method for early detection of cognitive aging. Also, the retrieval process for items from the same semantic category could lead to either similarity-based interference or facilitation effects. This aspect certainly presents an intriguing area for further exploration. However, this study was not primarily driven by the intention to make such a comparison, and the item count is insufficient for revisiting the analyses in this context. Future studies need to delve into this topic from diverse perspectives.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13534-023-00315-6.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jiae Kim and Jiyeon Lee. The first draft of the manuscript was written by Jiae Kim and Jiyeon Lee and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Research Council of Science & Technology(NST) grant by the Korea government (MSIT) (No. CAP21052-000; 2022R1A2C2005062) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2022R1I1A4063209).

Data Availability The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval The studies involving human participants were reviewed and approved by the Institutional Review Board of Ewha Womans University (EWIRB-202205-0032-01). The participants provided their written informed consent to participate in this study.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish The authors affirm that human research participants provided informed consent for publication of the images in Figs. 3a and b and 4a and b.

References

1. Hess E, Polt J. Pupil size in relation to mental activity during simple problem-solving. Science. 1964;143(3611):1190–2. https://doi.org/10.1126/science.143.3611.1190.
2. Kahneman D, Beatty J. Pupil diameter and load on memory. Science. 1966;154(3756):1583–5. https://doi.org/10.1126/science.154.3756.1583.
3. Granholm E, Asarnow RF, Sarkin AJ, Dykes KL. Pupillary responses index cognitive resource limitations. Psychophysiology. 1996;33(4):457–61. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x.
4. Cesana-Arlotti N, Varga B, Téglás E. The pupillometry of the possible: an investigation of infants' representation of alternative possibilities. Philos Trans R Soc B. 2022;377(1866):20210343. https://doi.org/10.1098/rstb.2021.0343.
5. Dietz J, Bradley MM, Okun MS, Bowers D. Emotion and ocular responses in Parkinson's disease. Neuropsychologia. 2011;49(12):3247–53. https://doi.org/10.1016/j.neuropsychologia.2011.07.029.

6. Ariel R, Castel AD. Eyes wide open: enhanced pupil dilation when selectively studying important information. Exp Brain Res. 2014;232:337–44. https://doi.org/10.1007/s00221-013-3744-5.
7. Heaver B, Hutton SB. Keeping an eye on the truth? Pupil size changes associated with recognition memory. Memory. 2011;19(4):398–405. https://doi.org/10.1080/09658211.2011.575788.
8. Kuchinsky SE, Ahlstrom JB, Vaden KI Jr, Cute SL, Humes LE, Dubno JR, et al. Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. Psychophysiology. 2013;50(1):23–34. https://doi.org/10.1111/j.1469-8986.2012.01477.x.
9. Schmidtke J. Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. Front Psychol. 2014;5:137. https://doi.org/10.3389/fpsyg.2014.00137.
10. Ellingsen DM, Wessberg J, Chelnokova O, Olausson H, Laeng B, Leknes S. In touch with your emotions: oxytocin and touch change social impressions while others' facial expressions can alter touch. Psychoneuroendocrinology. 2014;39:11–20. https://doi.org/10.1016/j.psyneuen.2013.09.017.
11. Baddeley A. Working memory. Science. 1992;255:556–9. https://doi.org/10.1126/science.1736359.
12. Baddeley AD, Hitch G. Working memory. In: Bower GA, editor. Psychology of learning and motivation. New York: Academic Press; 1974. pp. 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.
13. Banken JA. Clinical utility of considering digits forward and digits backward as separate components of the wechsler adult intelligence scale-revised. J Clin Psychol. 1985;41(5):686–91. https://doi.org/10.1002/1097-4679(198509)41:5%3C686::AID-JCLP2270410517%3E3.0.CO;2-D.
14. Ryan JJ, Lopez SJ, Paolo AM. Digit-span performance of persons 75–96 years of age: base rates and associations with selected demographic variables. Psychol Assess. 1996;8(3):324.
15. Suwartono C, Halim M, Hidajat L, Hendriks M, Kessels R. Development and reliability of the indonesian Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). Psychol. 2014;5:1611–19. https://doi.org/10.4236/psych.2014.514171.
16. Hester R, Kinsella G, Ong B. Effect of age on forward and backward span tasks. J Int Neuropsychol Soc. 2004;10(4):475–81. https://doi.org/10.1017/S1355617704104037.
17. Orsini A, Chiacchio L, Cinque M, Cocchiaro C, Schiappa O, Grossi D. Effects of age, education and sex on two tests of immediate memory: a study of normal subjects from 20 to 99 years of age. Percept Mot Skills. 1986;63(2):727–32. https://doi.org/10.2466/pms.1986.63.2.727.
18. Waters GS, Caplan D. Age, working memory, and on-line syntactic processing in sentence comprehension. Psychol Aging. 2001;16(1):128–44. https://doi.org/10.1037/0882-7974.16.1.128.
19. Johnson EL, Miller Singley AT, Peckham AD, Johnson SL, Bunge SA. Task-evoked pupillometry provides a window into the development of short-term memory capacity. Front Psychol. 2014;5:218. https://doi.org/10.3389/fpsyg.2014.00218.
20. Piquado T, Isaacowitz D, Wingfield A. Pupillometry as a measure of cognitive effort in younger and older adults. Psychophysiology. 2010;47(3):560–9. https://doi.org/10.1111/j.1469-8986.2009.00947.x.
21. Granholm EL, Panizzon MS, Elman JA, Jak AJ, Hauger RL, Bondi MW, et al. Pupillary responses as a biomarker of early risk for Alzheimer's disease. J Alzheimers Dis. 2017;56(4):1419–28. https://doi.org/10.3233/JAD-161078.
22. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol. 1999;56(3):303–8. https://doi.org/10.1001/archneur.56.3.303.
23. Martin N, Kohen F, Kalinyak-Fliszar M. A diagnostic battery to assess language and short-term memory deficits in aphasia, Clin Aphasiology Paper. 2008. http://aphasiology.pitt.edu/id/eprint/1944.
24. Sung JE. Performances on short-term and working memory tasks and their relationships to aphasia severity and auditory comprehension in normal elderly adults and people with aphasia. Commun Sci Disord. 2010;15(3):285–97.
25. Wingfield A, Stine EA, Lahar CJ, Aberdeen JS. Does the capacity of working memory change with age? Exp Aging Res. 1988;14(2):103–7. https://doi.org/10.1080/03610738808259731.
26. Christensen K, Multhaup K, Nordstrom S, Voss K. A cognitive battery for dementia: development and measurement characteristics. Psychol Assessment: J Clin Psychol. 1991;3(2):168. https://doi.org/10.1037/1040-3590.3.2.168.
27. Humes LE. What is normal hearing for older adults and can normal-hearing older adults benefit from hearing care intervention? Hear Rev. 2020;27(7):12–8.
28. Yeom TH, Park YS, Oh KJ, Kim JK, Lee YH. A manual for K-WAIS. Seoul: Korean Guidance Press; 1992. pp. 105–70.
29. Sung JE. The reliability and validity of short-term and Working Memory pointing tasks developed for clinical populations with Speech and Language Disorders. Commun Sci Disord. 2011;16(2):185–201.
30. Seo SK. Yonsei corpus of lexical frequency in korean vocabulary. Seoul: Yonsei University Institute of Language and Information Studies Press; 1998.
31. Govender A, King S. Using pupillometry to measure the cognitive load of synthetic speech. System. 2018;50:2838–42. https://doi.org/10.21437/Interspeech.2018-1174.
32. van Rij J, Hendriks P, van Rijn H, Baayen RH, Wood SN. Analyzing the Time Course of Pupillometric Data. Trends in Hearing. 2019;23. https://doi.org/10.1177/2331216519832483.
33. Mannaru P, Balasingam B, Pattipati K, Sibley C, Coyne JT. Performance evaluation of the gazepoint GP3 eye tracking device based on pupil dilation. Int. Conf. Augmented cognition. Switzerland AG: Springer Press; 2017. 166–75. https://doi.org/10.1007/978-3-319-58628-1.
34. Caplan D, Waters G. The relationship between age, processing speed, working memory capacity, and language comprehension. Memory. 2005;13(3–4):403–13. https://doi.org/10.1080/09658210344000459.
35. Salthouse TA. The aging of working memory. Neuropsychol. 1994;8(4):535–43. https://doi.org/10.1037/0894-4105.8.4.535.
36. Cowan N, Elliott EM, Saults JS, Morey CC, Mattox S, Hismjatullina A, et al. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. Cogn Psychol. 2005;51(1):42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001.
37. Craik FIM, Jennings JM. Human memory. In: Craik FIM, Salthouse TA, editors. The handbook of aging and cognition. Lawrence Erlbaum Associates Inc; 1992. pp. 51–110.
38. Light LL. Memory and aging: four hypotheses in search of data. Annu Rev Psychol. 1991;42:333–76. https://doi.org/10.1146/annurev.ps.42.020191.002001.
39. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev. 1996;103(3):403–28. https://doi.org/10.1037/0033-295X.103.3.403.
40. Eom BR, Oh SJ, Sung JE. Task-specific effects on performance of working memory measures and their relationship to aphasia severity in people with aphasia. J Speech lang hear Disord. 2016;25(2):113–22.
41. Jones G, Macken B. Questioning short-term memory and its measurement: why digit-span measures long-term associative learning. Cognition. 2015;144:1–13. https://doi.org/10.1016/j.cognition.2015.07.009.

42. Just MA, Carpenter PA. A capacity theory of comprehension: individual differences in working memory. Psychol rev. 1992;99(1):122. https://doi.org/10.1037/0033-295X.99.1.122.