Task repetition influence on pupil response during encoding of auditory information in normal-hearing adults

MISEUNG KOO[1], MYUNG-WHAN SUH[1,2], JUN HO LEE[1,2], SEUNG-HA OH[1,2], AND MOO KYUN PARK[1,2*]

[1] *Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Hospital, Seoul, Korea*

[2] *Sensory Organ Research Institute, Seoul National University Medical Research Center, Seoul, Korea*

Although numerous behavioural measures to estimate listening effort have been developed in recent years using free recall or dual-task paradigms, relatively little is known about physiological measures, such as pupil dilation, in response to cognitively demanding tasks. This study used a repeated-measure experimental design and aimed to investigate the cognitive resource allocation process of spoken words in an immediate free recall paradigm. Here, ten adults with normal hearing (NH) attended 2 days of trials with 14 trials per day. The listeners heard four-speaker babble noise along with seven sentences and then tried to remember the first words of all seven sentences. Recall performance on the first day only showed a significant serial position effect ($p < 0.05$). With increasing memory load imposed by the subsequent recall task, baseline pupil size significantly enlarged ($p < 0.01$), and the PPDs significantly decreased ($p < 0.01$) during the encoding process, implying that a gradual increase in resources allocated to memory capacity corresponded to a decline in resources allocated to listening. Real-time allocation of cognitive resources during the encoding of spoken words can be monitored independently by the analysis of pupil dilation averaged over multiple trials.

INTRODUCTION

Hearing-impaired (HI) listeners may have to devote more effort to perceiving speech under adverse listening conditions (Kahneman, 1973). Interestingly, effortful listening in everyday conversation where the speech is fully audible and intelligible has been reported (Lunner *et al.*, 2016; Pichora-Fuller *et al.*, 2016), implying that individuals devote different amounts of effort to facilitate understanding even though behavioural performances may not differ. Listening effort, defined in the framework for understanding effortful listening (FUEL) as “deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task” (Pichora-Fuller *et al.*, 2016), has become a subject of increasing interest in cognitive hearing science (Rudner *et al.*, 2014). Numerous assessment tools have been developed to assess individual cognitive spare capacity in relation to varying effort in listening, including the sentence-final word identification and recall test (SWIR) (Ng *et al.*, 2013; Ng *et al.*, 2015) that inspired our research. However, behavioural measures

*Corresponding author: aseptic@snu.ac.kr

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

alone might not fully describe the quantitative change in listening effort during encoding of auditory stimuli.

Pupillometry has recently gained considerable attention as the most promising physiological measure underlying cognitive processing in response to hearing-related tasks (Pichora-Fuller *et al.*, 2016; Zekveld *et al.*, 2018). Studies on speech recognition in noise frequently utilise peak pupil dilation (PPD) to quantify changes in listening effort (Ohlenforst *et al.*, 2017). In addition, a greater pre-stimulus baseline, also known as tonic or baseline pupil size (BS), was observed in participants with higher working memory capacity (WMC) (Heitz *et al.*, 2008; Tsukahara *et al.*, 2016). We hypothesised that BS would vary not only from subject to subject, reflecting variations in individual WMC, but also between sentences in a trial, with changes in memory demand imposed by the recall task. This study had the following aims:

1. To investigate whether a repeated-measure experimental design would have a favourable impact on either behavioural assessment or pupillometric data in response to a cognitive task;
2. To monitor task-evoked changes in pupil size in response to the recall task to determine whether real-time cognitive resource allocation can be detected based on pupil size during the encoding of auditory stimuli.

MATERIALS AND METHODS

Twelve fluent Korean speakers (mean age 24.6 years, range 22–29 years, eight males) with NH were initially recruited. All participants reported normal or corrected-to-normal vision as well as bilateral NH and attended two visits (days 1 and 2) with at least a 3-week interval to avoid any learning effect (Ohlenforst *et al.*, 2018; Simonsen *et al.*, 2016). They were told to refrain from caffeine consumption for at least 6 h before each visit. After their first visit, two of the participants were excluded from data collection due to large amounts of missing data.

Stimuli

Fourteen seven-sentence lists from the Korean version of the hearing in noise test (HINT) (Moon *et al.*, 2005) were selected for the SWIR in accordance with the published study protocol (Lunner *et al.*, 2016; Ng *et al.*, 2013; Ng *et al.*, 2015). Target speech, spoken by a male speaker, was presented at 65 dB SPL along with four-talker babble noise (two males and two females) starting 2 s before sentence onset and ending 2 s after sentence offset (Fig. 1). To evaluate serial-position effects, sentences in each set of seven sentences were allocated as follows: the first and second sentences to the primacy, third to fifth sentences to the asymptote, and sixth to seventh sentences to the recency position.

Procedure

During the first visit, the HINT speech reception thresholds (SRTs), speech and noise from 0° and at 80% correct performance, were obtained from individuals using a published HINT procedure (Hallgren *et al.*, 2006). Subsequently, during SWIR training with four practices, the $SRT_{80\%}$ was tuned to reach the signal-to-noise ratio (SNR) 95% correct performance depending on repetition performance of the first words of each sentence (identification task), as described in Ng et al. (2015), although listeners were instructed to complete both tasks: identification and free recall. The recall phase began with the presentation of a 0.5 s beep sound, and participants were prompted to report the first words in any order as many as possible (recall task). In the following blocks comprising 10 trials of a recall task in the SWIR and pupil diameter recording, participants were not required to give any verbal response before the beep in order to prevent rehearsal of to-be-remembered items that might potentially influence the subsequent recall performance. At the second visit, participants repeated the SWIR training and SWIR with pupil data recording while maintaining the SNRs obtained from the first visit (Fig. 1).

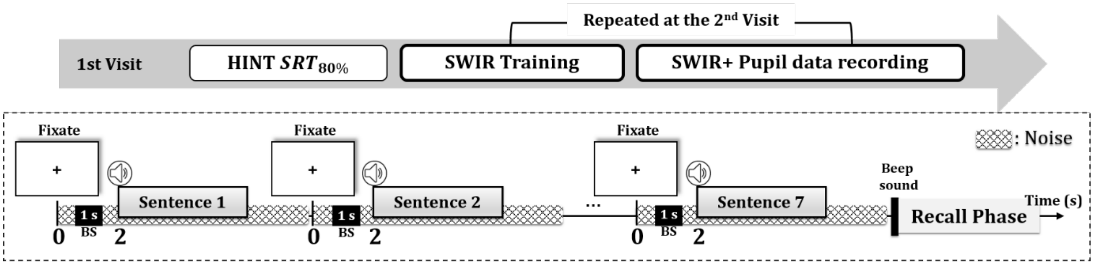

Fig. 1: Upper panel: Three experimental sessions used in this study. Lower panel: Encoding and recall phases of a trial in SWIR. The inter-stimulus interval between sentence offset and onset of the following sentence was longer than 4 s. "0" represents the start of four-talker babble noise used to calculate the peak pupil latency.

Pupil diameter was recorded using a wearable eye-tracking headset (Pupil Labs, Germany) with 200 Hz binocular eye cameras positioned in front of the eyes and using Matlab software (Release 2018a) provided by Oticon Medical A/S, Smørum, Denmark. The raw pupil diameter data were pre-processed to remove samples with blink artifacts or dilation speed outliers using median absolute deviation method, as described in Kret *et al.* (2019). Pupil diameter values greater or lower than the median ± 2.5 times the standard deviation of the remaining data were defined as blink. We also applied divisive baseline correction (proportion change: corrected pupil size = pupil size/baseline) to the pre-processed data at the end of normalization. Room illumination was provided by two LED lights positioned on the ceiling of the testing booth and varied depending on the dynamic range of the participant's pupil size to prevent floor and ceiling effects, and was individually adjusted to the pupil-size midpoint, from dim (~30 lux) to bright (~230 lux) prior to data collection, with an average illuminance of 110 lux.

Statistical analyses

A nonparametric repeated-measures analysis of variance (RM ANOVA) was used to analyse recall performance with two within-subjects factors: word position (primacy, asymptote, recency) and visit (day 1 and 2) because the recall score was a discrete variable rather than a continuous one that follows a normal distribution.

Nonparametric analysis of repeated data was performed with R (nparLD package) (Brunner *et al.*, 2002; Noguchi *et al.*, 2012).

Statistical analyses of the pupil data collected from the encoding phase were performed using SPSS software, version 25 (Chicago, IL, USA). A linear mixed model (LMM) was employed to examine the data because of its ability to handle missing values due to the large number of blinks and to statistically compare the fixed effects of stimulus presentation order and visit on BS, PPD, and peak pupil latency (time interval between sentence onset and PPD). The average pupil size during the 1 s pre-stimulus period served as BS. Post hoc analyses with Bonferroni correction were used to adjust for multiple comparisons. Eye was not included as a fixed effect because, in our preliminary experiments, no significant influence of eye on pupil response was observed. A p-value <0.05 was considered significant. After data selection, we identified 298 invalid pupil traces out of 2,800 pupil traces, recorded per sentence, due to missing PPD values for containing either more than 25% blinks or erroneous recording. We measured 35.7 pupil recordings per participant on average, regardless of eye position.

RESULTS

Recall performance

As depicted in Fig. 2, the nonparametric RM ANOVA results revealed a significant interaction between word position and visit day, indicating that a significant serial position effect was found in recall performance on the first day only ($p = 0.0274$). Post hoc analyses with Bonferroni correction showed significantly better performance for early (primacy) than for late (recency) items in the list ($p < 0.0167$). This pattern, however, was not seen in recall performance on the second day. No other significant main effects or interactions were observed.

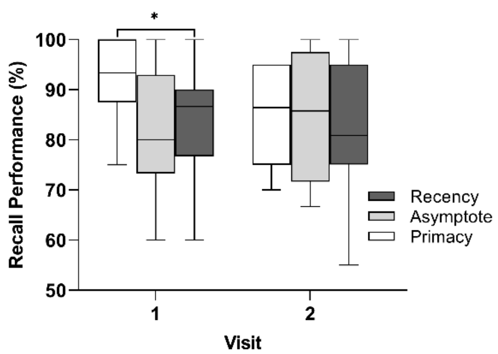

Fig. 2: Results of participants' recall scores as a function of word position (primacy, asymptote, and recency) and visit (day 1 and 2).

BS, PPD, and peak pupil latency during memory encoding of spoken words

The LMM results for BS revealed significant main effects of visit ($F = 67.62$, $p < 0.001$) and stimulus presentation order ($F = 41.63$, $p < 0.001$), in addition to a significant interaction between these variables (F = 3.48, p = 0.002). Compared with BS for the first item, listeners' pupil size increased progressively with increasing memory load in preparation for the subsequent recall task (Fig. 3a). The incremental increase in recall performance was greater on the second day than that on the first day (Fig. 3b).

The LMM results for the PPD revealed significant main effects of stimulus presentation order (F = 13.10, p < 0.001) and visit day (F = 8.67, p = 0.003) in addition to a significant interaction between these variables (F = 2.52, p = 0.02). There was a progressive decline in PPD as the number of words to be remembered increased (Fig. 3c), and the PPD on the second day was significantly greater than that on the first day (Fig. 3d).

The LMM for peak pupil latency revealed significant main effects of visit day (F = 46.08, p < 0.01) and stimulus presentation order (F = 4.58, p < 0.01); however, no significant interaction was observed. Post hoc analyses using Bonferroni correction showed that the latency was significantly shorter for the first items than for the middle items, i.e., the third (p = 0.027), fourth (p < 0.001), and sixth sentences (p = 0.004) (Fig. 3e). The latency on day 1 was significantly shorter than that on day 2 (Fig. 3f).

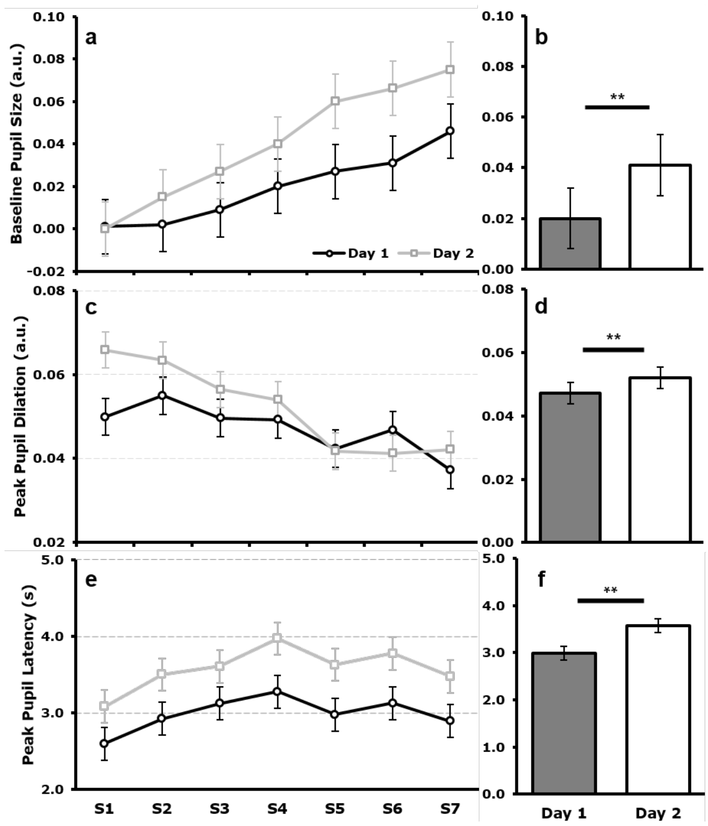

Fig. 3: Mean ± 1 SE of BS (a, b), PPD (c, d), and peak pupil latency (e, f) during memory encoding as a function of stimulus presentation order (S1 to S7) and visit.

DISCUSSION

Our findings agree with earlier literature showing that single-trial analysis is insufficient to provide a reliable interpretation of pupillometry data (Winn *et al.*, 2018). This experiment, which attempted to quantify real-time allocation of cognitive resources during encoding of auditory inputs, also expanded the findings of previous studies (Lunner *et al.*, 2016; Ng *et al.*, 2013) by adding time-locked eye tracking analysis of the memory-encoding period followed by the free recall task.

Our most important finding was that the BS tended to increase; while the PPD tended to decrease during the phase when participants were encoding words heard against a competing speech background (see Fig. 3). These incremental changes are closely associated with increasing memory demands as listeners strive to memorise the items for a subsequent recall task. BS, recorded sentence-by-sentence during the pre-stimulus period, seemed to be an effective indicator of the changes in memory load and the ability to store a sequence of auditory items, in line with previous findings (Tsukahara *et al.*, 2016). There are a number of different perspectives on the role of BS. So far most studies have used a trial-based analysis, including Gilzenrat *et al.* (2010) who linked increases in BS to reduced task utility and disengagement from a given task with the purpose to demonstrate an inversed relationship between BS and task-evoked pupil dilations, predicted by the adaptive-gain theory (Aston-Jones *et al.*, 2005). Since the listening condition used in this study was relatively easy for NH listeners, it is predetermined that the recall task might be more and more rewarding for them to explore, relative to the identification (control) task, thereby producing BS that progressively enlarged over the course of a trial.

The longest peak latency for the middle items seemed to associate item difficulty or high processing load on items in the asymptote position (primacy effect). Similarly, Koelewijn *et al.* (2014) found longer latency to PPD in the dual-sentence condition compared to the single-sentence condition. However, recall performance in asymptote varied between the visits, and indeed it was the highest in the second visit. This consistent pattern of peak latency needs to be incorporated with recall performance, which was not examined in our study, in further study. In addition, researchers when designing multiple conditions and experiments repeated on different days may need to consider the effect of test-day on pupil response because a statistically significant increase in BS, PPD, and peak latency was found in our study. We would recommend a minimum of 10 repetitions to be carried out for each experimental condition excluding the first few trials (i.e. familiarization training).

One weakness of our study is that the use of contact lenses was not restricted due to difficulties in recruiting suitable subjects who had both NH and normal, uncorrected vision. Rather some participants were allowed to use them to avoid any unnecessary fatigue in maintaining eye fixation (not the focus of our study). To increase data accuracy and reduce variability, our findings should be confirmed in further studies using a large numbers of subjects and examining between-subject factors such as age, hearing status (e.g. HI listeners), and cognitive ability because listening effort studies may help future hearing rehabilitation practices and approaches (Richmond et al., 2011). Since task-evoked pupil dilation is apparently much smaller than other pupillary reflexes, we focused on NH young adults who reportedly exhibit greater changes in pupil response than older age and HI groups (Winn *et al.*, 2018). Moreover, an adequate normalisation or analysis method (e.g. growth curve analysis) should be employed to minimise possible individual variance.

In conclusion, pupillometry can be an independent indicator for monitoring online resource allocation in a free recall paradigm where a repeated-measure design is adopted. Although we could not explore inter-individual variance in cognitive processing using the analysis of pupil dilation in this study, pupillometry was able to detect the ongoing changes during the memory-encoding phase while behavioural assessments, measured offline, could not provide such information.

REFERENCES

Aston-Jones, G., and Cohen, J. D. (**2005**). “An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance,” Annu. Rev. Neurosci., **28**, 403-450. doi:10.1146/annurev.neuro.28.061604.135709.

Brunner, E., Domhof, S., and Langer, F. (**2002**). “Nonparametric analysis of longitudinal data in factorial experiments,” New York, NY: J. Wiley.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. (**2010**). “Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function,” Cogn. Affect. Behav. Neurosci., **10**(2), 252-269. doi:10.3758/CABN.10.2.252.

Hallgren, M., Larsby, B., and Arlinger, S. (**2006**). “A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition,” Int. J. Audiol., **45**(4), 227-237. doi:10.1080/14992020500429583.

Heitz, R. P., Schrock, J. C., Payne, T. W., and Engle, R. W. (**2008**). “Effects of incentive on working memory capacity: behavioral and pupillometric data,” Psychophysiology, **45**(1), 119-129. doi:10.1111/j.1469-8986.2007.00605.x.

Kahneman, D. (**1973**). “Attention and effort,” Englewood Cliffs, N.J., Prentice-Hall.

Kret, M. E., and Sjak-Shie, E. E. (**2019**). “Preprocessing pupil size data: Guidelines and code,” Behav. Res. Methods, **51**(3), 1336-1342. doi:10.3758/s13428-018-1075-y.

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., and Kramer, S. E. (**2014**). “The pupil response is sensitive to divided attention during speech processing,” Hear. Res., **312**, 114-120. doi:10.1016/j.heares.2014.03.010.

Lunner, T., Rudner, M., Rosenbom, T., Agren, J., and Ng, E. H. (**2016**). “Using Speech Recall in Hearing Aid Fitting and Outcome Evaluation Under Ecological Test Conditions,” Ear Hear., **37**(Suppl 1), 145S-154S. doi: 110.1097/AUD.0000000000000294.

Moon, S. K., Mun, H. A., Jung, H. K., Soli, S. D., Lee, J. H., and Park, K. (**2005**). “Development of Sentences for Korean Hearing in Noise Test (KHINT),” Korean J. Otolaryngol., **48**, 724-728. doi:10.1097/AUD.0b013e31803154d0.

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., and Ronnberg, J. (**2013**). “Effects of noise and working memory capacity on memory processing of speech for hearing-aid users," Int. J. Audiol., **52**(7), 433-441. doi:10.3109/14992027.2013.776181.

Ng, E. H., Rudner, M., Lunner, T., and Ronnberg, J. (**2015**). "Noise reduction improves memory for target language speech in competing native but not foreign language speech," Ear Hear., **36**(1), 82-91. doi:10.1097/AUD.0000000000000080.

Noguchi, K., Gel, Y. R., Brunner, E., and Konietschke, F. (**2012**). "nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments," Journal of Statistical Software, **50**(12), 1-23.

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (**2018**). "Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response," Hear. Res., **365**, 90-99. doi:10.1016/j.heares.2018.05.003.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., . . . Kramer, S. E. (**2017**). "Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation," Hear. Res., **351**:68-79. doi: 10.1016/j.heares.2017.05.012. Epub 2017 May 1025.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., . . . Wingfield, A. (**2016**). "Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)," Ear Hear., **37**(Suppl 1), 5S-27S. doi: 10.1097/AUD.0000000000000312.

Richmond, L. L., Morrison, A. B., Chein, J. M., and Olson, I. R. (**2011**). "Working memory training and transfer in older adults," Psychol. Aging, **26**(4), 813-822. doi:10.1037/a0023631.

Rudner, M., and Lunner, T. (**2014**). "Cognitive spare capacity and speech communication: a narrative overview," BioMed Res. Int., 2014, 869726. doi:10.1155/2014/869726.

Simonsen, L. B., Hietkamp, R. K., and Bramsløw, L. (**2016**). "Learning effects of repeated exposure to Hearing In Noise Test," Paper presented at the Annual Conference of the British Society of Audiology, Coventry, UK.

Tsukahara, J. S., Harrison, T. L., and Engle, R. W. (**2016**). "The relationship between baseline pupil size and intelligence," Cogn. Psychol., **91**, 109-123. doi:10.1016/j.cogpsych.2016.10.001.

Winn, M. B., Wendt, D., Koelewijn, T., and Kuchinsky, S. E. (**2018**). "Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started," Trends Hear., **22**, 1-32. doi:10.1177/2331216518800869.

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (**2018**). "The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge," Trends Hear., **22**, 2331216518777174. doi:10.1177/2331216518777174.