Original Article

QJEP

Quarterly Journal of Experimental Psychology
2021, Vol. 74(4) 705–715

DOI: 10.1177/1747021820972255
qjep.sagepub.com

Common principles underlie the fluctuation of auditory and visual sustained attention

Hiroki Terashima[1], Ken Kihara[2], Jun I Kawahara[3] and Hirohito M Kondo[1,4]

Abstract

Sustained attention plays an important role in adaptive behaviours in everyday activities. As previous studies have mostly focused on vision, and attentional resources have been thought to be specific to sensory modalities, it is still unclear how mechanisms of attentional fluctuations overlap between visual and auditory modalities. To reduce the effects of sudden stimulus onsets, we developed a new gradual-onset continuous performance task (gradCPT) in the auditory domain and compared dynamic fluctuation of sustained attention in vision and audition. In the auditory gradCPT, participants were instructed to listen to a stream of narrations and judge the gender of each narration. In the visual gradCPT, they were asked to observe a stream of scenery images and indicate whether the scene was a city or mountain. Our within-individual comparison revealed that auditory and visual attention are similar in terms of the false alarm rate and dynamic properties including fluctuation frequency. Absolute timescales of the fluctuation in the two modalities were comparable, notwithstanding the difference in stimulus onset asynchrony. The results suggest that fluctuations of visual and auditory attention are underpinned by common principles and support models with a more central, modality-general controller.

Keywords

Sustained attention; attentional fluctuation; gradual-onset continuous performance task (gradCPT); hearing; vision

Received: 5 November 2019; revised: 12 October 2020; accepted: 14 October 2020

Introduction

Selective attention plays an important role in identifying transient, task-relevant information in a complex scene (Moore & Zirnsak, 2017; Nobre, 2018). Aside from functions for selective aspects, sustaining attention over time is critical for adaptive behaviours in everyday activities, such as driving, reading, and listening to a lecture (Esterman et al., 2013), and for some people, such as telemarketers and air traffic controllers, the ability to maintain an appropriate level of auditory and visual attention is a required professional skill (Imbert et al., 2014; Kim et al., 2018). However, it is effortful to continue paying attention to a currently engaging task (Warm et al., 2008), because the vulnerability of the focused state of attention to exogenous external events and endogenous internal factors can cause it to momentarily fluctuate or even be lost.

Such temporal vulnerabilities of attention hint at one of the fundamental questions in cognitive psychology: how are attentional resources dynamically allocated across mental processes? Sustained attention has been typically modelled either as resource depletion (overload) or as mindlessness (underload) (Fortenbaugh et al., 2017; Thomson et al., 2015). The two views were recently integrated into a resource-control model (Thomson et al., 2015), in which a controller (a higher level system) allocates attentional resources to the task and the amount of

[1]NTT Communication Science Laboratories, Nippon Telegraph and Telephone, Atsugi, Japan
[2]Department of Information Technology and Human Factors, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
[3]Department of Psychology, Hokkaido University, Sapporo, Japan
[4]School of Psychology, Chukyo University, Nagoya, Japan

Corresponding author:
Hiroki Terashima, Human Information Science Laboratory, NTT Communication Science Laboratories, Nippon Telegraph and Telephone, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan.
Email: hiroki.terashima.cs@hco.ntt.co.jp

the allocation tends to decrease in later trials as more resources are allotted for mind-wandering (the default target of allocation). As this line of discussion has been based mainly on studies in the modality of vision and given that temporal aspects of visual and auditory attention are not necessarily comparable (Zalta et al., 2020), it is still unclear whether the model can be generalised to other sensory modalities such as audition.

Predictions of how much the fluctuation of sustained attention is modality-general depend on theories. From viewpoints focusing on attentional resources, the resources in vision and audition have been thought to be distinct (Alais et al., 2006; Duncan et al., 1997; Larsen et al., 2003), in accordance with the general multiple-resource framework in the literature of human attention (Baddeley, 2012; Wickens, 2008). In line with the framework, a recent study revealed distinct natural rhythms for auditory and visual temporal attention (Zalta et al., 2020). Simple resource-based models thus do not necessarily predict similarity between visual and auditory sustained attention. On the contrary, a more central, modality-general mechanism has been shown to play a central role in some attentional phenomena (Jolicoeur, 1999; Lehnert & Zimmer, 2008; Saults & Cowan, 2007). For sustained attention, the resource-control model explained the attentional fluctuation as the variability of resource allocation by the controller (Thomson et al., 2015). Because the controller is an integrated higher level system, the model predicts that the characteristics of the attentional fluctuation are not specific to vision but rather shared with audition. As the fluctuation of sustained attention is inherently dynamic, we should directly compare the dynamic nature of the resource allocation process in auditory and visual modalities. However, we lack a proper experimental paradigm for the between-modality comparisons in terms of fluctuation dynamics.

A continuous performance task (CPT) has been one of the most useful measures for assessing the level of sustained attention or vigilance. In such a task, participants respond to frequent non-targets and withhold their responses to infrequent targets, which implicitly calls for maintaining an attentional level. Researchers have found that CPT performance depends on an interaction of three factors: task parameters, participant's personality traits, and environmental conditions (Ballard, 2001). The present study focused on task parameters. Previous studies have shown that a critical parameter is the speed of stimulus presentation: faster event rates in CPTs lead to fewer correct responses and longer reaction times (RTs) (Parasuraman & Giambra, 1991), and the CPT performance is associated with everyday cognitive failures (McCrae, 2007). However, there were methodological limitations in those studies. Trials in these tasks were accompanied by sudden stimulus onsets, which may reduce demands on the maintenance of endogenous attention (Sturm & Willmes, 2001). When sudden-onset visual cues were presented before target stimuli, perceptual sensitivity was enhanced in a vigilance task (MacLean et al., 2009). Another issue is a lack of analyses that can assess trial-to-trial, within-individual dynamics of sustained attention (Esterman et al., 2013). Performance measures are limited to temporally summarised scores such as mean accuracy.

Such limitations were shared by a few CPT studies that compared auditory and visual sustained attention, which may explain their mixed results. The auditory CPT appeared to be more difficult than the visual CPT (Baker et al., 1995), but the performance difference decreased with advancing age (Aylward et al., 2002). Visual vigilance tasks appeared more stressful for adults than auditory vigilance tasks (Galinsky et al., 1993; Szalma et al., 2004). A previous study used discrete spoken digits for the auditory task and visually presented digits for the visual task (Seli et al., 2012). The study reported significant between-modality correlations for the false alarm (FA) rate and reaction time (RT) variability. However, again, performance measured in these studies may be affected by sudden stimulus onsets, and the dynamics of fluctuation was not considered. To test whether the correlations are genuinely diagnostic of common mechanisms of sustained attention or not, correlations of the indices and dynamic properties should be investigated under conditions where sudden stimulus onsets are removed.

To reduce the attentional capture caused by the sudden image onsets of conventional CPTs and elucidate temporal dynamics of attention, a gradual-onset CPT (gradCPT) has been developed (Esterman et al., 2013). In the gradCPT, they presented visual images that gradually changed from one to the next, and participants judged whether the stimulus was a scene of a city or mountain. The results revealed a tight link between FA rates and RT variability. Their novel analysis enables us to track the dynamics of attentional fluctuations via a time series of RTs. Although the gradCPT has been applied only in the visual modality so far, it should also be possible for the gradCPT to remove the effects of sudden stimulus onsets on auditory-visual comparison.

The present study developed an auditory analogy to the visual gradCPT, in which narrations gradually changed from one to the next without sudden onsets, and participants judged whether the stimulus was a male or female voice. In Experiment 1, we investigated the degree to which the auditory and visual gradCPT performance varied with stimulus onset asynchrony (SOA) and chose a longer SOA for the auditory gradCPT to equalise task difficulty. In Experiment 2, we identified the fluctuation of sustained attention in both modalities and found similarities in task performance and fluctuation frequencies. The results suggest that some principles underlie the dynamic resource allocation process in sustained auditory and visual attention.

Method

Participants

A total of 24 college students participated in Experiment 1 (11 males and 13 females; mean age 19.9 years, range 18–21 years). Another 29 students participated in Experiment 2 (14 males and 15 females; mean age 20.7 years, range 18–25 years). We based our sample size on the previous literature (Esterman et al., 2013) and a power analysis. According to a power analysis with a power of 0.8 (α-level = .05), we required 28 participants for Experiment 1 to detect main effects and interactions in analysis of variance (ANOVA; $f = 0.25$, F-test) and 23 participants for Experiment 2 to detect between-modality correlations ($r = .5$; bivariate normal model). The number of participants for Experiment 1 was less than originally planned due to unexpected failures in data acquisition (two participants quit before the end; another two did not reach the hit rate of 50%). All participants were right-handed with normal hearing and normal or corrected-to-normal vision. None had any history of neurological or psychiatric disorders. This study was approved by the Ethics Committees of Chukyo University and Hokkaido University (approval nos. RS17-020 and 28-2) and was carried out under the Ethical Guidelines for Medical and Health Research Involving Human Subjects. Participants gave written informed consent after the procedures had been fully explained to them.

Behavioural tasks

For the auditory gradCPT, the stimuli consisted of narrations spoken by 10 males and 10 females. The stimuli were randomly presented with males (90%) and females (10%), without allowing identical narrations to repeat in consecutive trials. The narrations gradually changed from one to the next, using rising and falling sinusoidal ramps (Figure 1a). All narrations were normalised by the root mean square level, and the presentation level was adjusted to a comfortable listening level. The stimuli were delivered through Sennheiser HD 599 headphones. The narrations were chosen from a narrative database from a variety of languages (International Phonetic Association, 1999) so that the languages differed from the native language of the participants. Thus, the participants judged the voice stream's gender using acoustic clues of the stimuli, and not semantic clues.

The visual gradCPT was conducted in accordance with a previous study (Esterman et al., 2013). All stimuli were round, greyscale photographs that contained 10 city scenes and 10 mountain scenes. The stimuli were randomly presented with city (90%) and mountain (10%), without allowing identical scenes to repeat in consecutive trials. The scenes gradually changed from one to the next, using a linear pixel-by-pixel interpolation (Figure 1b). Images were subtended 4.5° of visual angle at 57 cm of viewing distance.

In Experiment 1, participants were randomly assigned to either the auditory or visual gradCPT. Each gradCPT consisted of three 8-min conditions. For each condition, the stimulus duration and SOAs were fixed as 1,600/800, 2,400/1,200, or 3,200/1,600 (auditory/visual) ms. The order of conditions was randomised across participants. In Experiment 2, participants performed both the auditory and visual gradCPTs. From the perspective of task difficulty measured in Experiment 1, the stimulus duration and SOA were 3,200 and 1,600 ms for the auditory gradCPT and 1,600 and 800 ms for the visual gradCPT. The order of the tasks was randomised across participants. The stimulus presentation and data collection were controlled by a PC with MATLAB and Psychtoolbox-3 (Brainard, 1997; Pelli, 1997). Participants were instructed to press a key for each male narration or each city scene and withhold responses to female narrations or a mountain scene. They responded to frequent targets as quickly and accurately as possible. Given that the next stimulus would replace the current stimulus within the SOA, a response deadline was implicit in the tasks.

Data analyses

Computation of RTs. We defined an RT to each stimulus as the relative time from the stimulus onset (Esterman et al., 2013). First, we set a time window into which typical RTs fell. For a stimulus, the window started after 70% had appeared and ended after 40% had disappeared in the disappearing phase. Any key press in the time window was assigned to the current stimulus. Next, when an ambiguous key press did not fall into the time window, it was considered as a response to an adjacent stimulus if there was no response to the stimulus. If both adjacent stimuli had no responses, we assigned the response to the closer stimulus. If one of them was a No-Go trial, we assigned the response to the other Go trial. Finally, if multiple key presses occurred in a single trial, we used the shortest RT for analysis.

Fluctuation of sustained attention. In Experiment 2, we used 8-min time-series data of RTs to quantify the fluctuation of sustained attention for each participant. RTs that were too quick and too slow could be considered as signatures of the lack of attention. We focused on the variance time course (VTC) of RT z-scores (Esterman et al., 2013). The VTCs are time series of the absolute values of z-scored RTs (not raw RTs) in each gradCPT. A VTC value between zero and one means that the corresponding RT has a typical (stable) value (within one SD). Before computing the VTCs, we interpolated RTs for the trial without responses using RTs of adjacent trials. A Gaussian kernel at the full-width at half-maximum of 7 s was applied to smooth the VTCs (Figure 2).

Frequency of attentional fluctuations. To quantify how sustained attention fluctuates, we performed a frequency analysis for

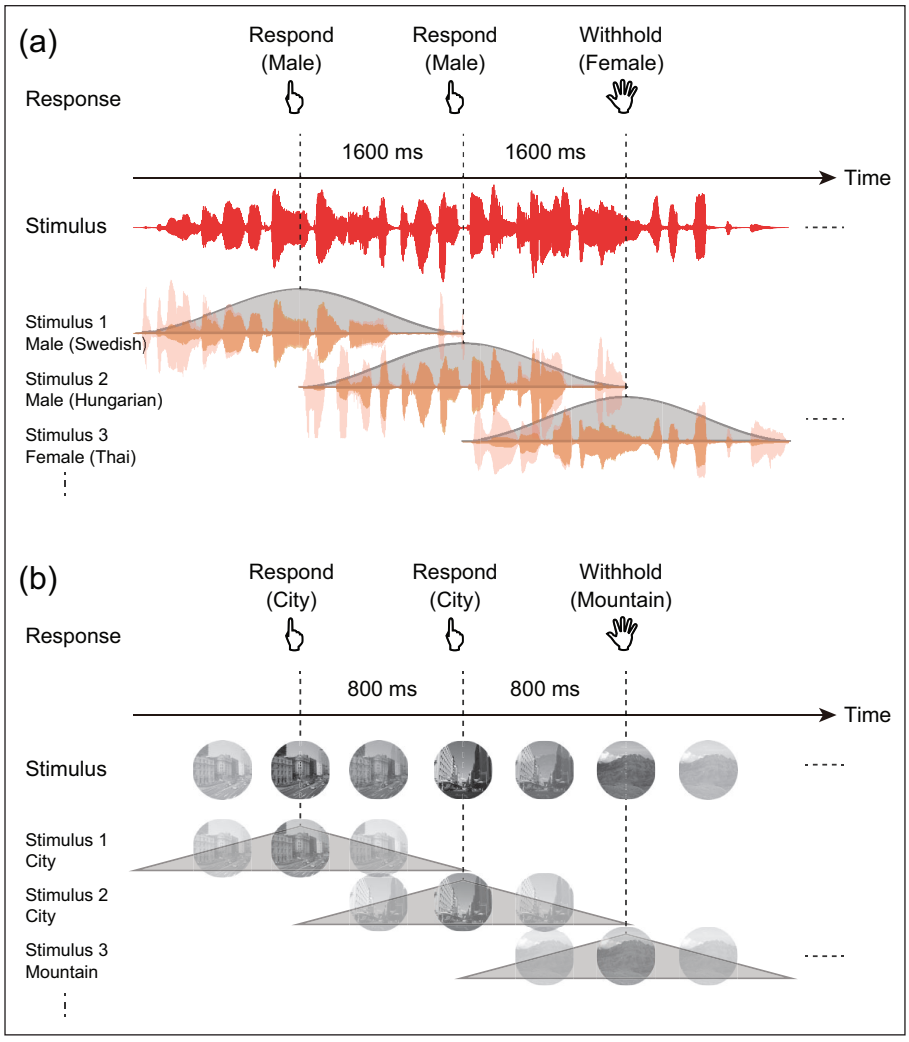

Figure 1. Schematic representation of the auditory and visual gradCPTs. (a and b) The presented sound (red) consists of stimuli that were consecutively presented without sudden onsets. Each voice narration (pale orange) in the auditory gradCPT was preprocessed with a sine bump (orange), similar to image preprocessing in the visual gradCPT.

VTCs. We applied a discrete Fourier analysis to a single VTC and obtained a frequency spectrum. We defined the VTC fluctuation frequency of each task as the frequency that has the largest power in the frequency range above 0.0025 Hz. To examine cognitive demands on sustained attention, we computed the time-dependent changes in the FA rate and RT coefficient of variation (CV = *SD*/average). A sliding window of 2-min width was used to obtain the time series of their values. All trials of the task were used to compute the representative value for each task.

Statistical analysis. In Experiment 1, we computed the sensitivity (d') and RTs for each condition and performed a mixed-design analysis of variance ANOVA. The Šidák correction was used for post hoc comparisons (α-level = .05). We assessed the speed–accuracy tradeoff by performing an ordinary least squares (OLS) regression for d' explained by RTs. In Experiment 2, we calculated the FA rate, hit rate, normalised criterion (C'), and RT CV as well as d' and RTs. The Jonckheere–Terpstra test was used as a trend test (two-tailed).

Our planned statistical analyses were the ANOVA, the OLS regression, between-modality correlation analyses of d', FA rate, hit rate, and VTC fluctuation frequency, within-modality correlation analyses between FA rate and RT CV, and trend tests for FA rate and RT CV. Statistical analyses were carried out with IBM SPSS Statistics (version 23) and R (version 3.1.2).

Results

Experiment 1: effects of SOAs on gradCPT performance

To choose an appropriate SOA for the auditory gradCPT, we performed a 2 (task types) × 3 (SOAs) ANOVA on task accuracy and RTs (Figure 3). The results showed that d' ($M \pm SD$) was greater for the visual gradCPT (3.50 ± 0.25) than for the auditory gradCPT (2.43 ± 0.25):

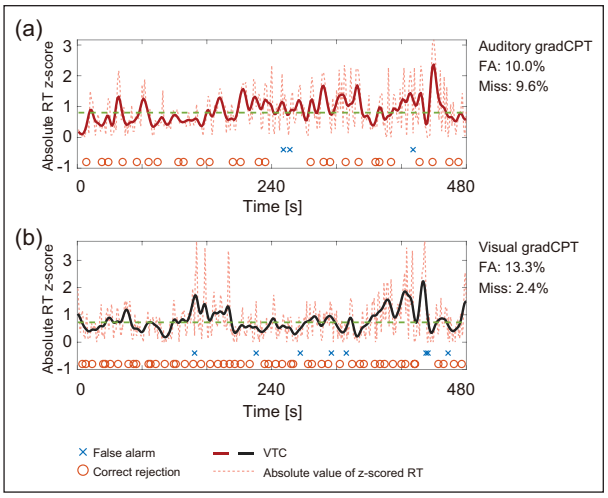

Figure 2. Time courses of the auditory and visual gradCPTs from a representative participant. (a and b) The reflected time course (bold line) of sustained attention. It was derived from z-scored RTs (dotted line) and smoothed. A larger value indicates a more variable state. The horizontal dashed line shows the median value of the VTC. The data are derived from Experiment 2.
VTC: variance time course.

$F(1, 22) = 9.20$, $\eta_p^2 = 0.30$, $p = .006$. For longer SOAs, d' improved: 2.24 ± 0.16 for 800 ms SOA, 3.19 ± 0.21 for 1,200 ms SOA, and 3.47 ± 0.21 for 1,600 ms SOA; $F(2, 44) = 38.96$, $\eta_p^2 = 0.64$, $p < .001$. The interaction between task types and SOAs was also significant: $F(2, 44) = 4.68$, $\eta_p^2 = 0.18$, $p = .014$. Next, a 2 × 3 ANOVA revealed that the mean RT was faster for the visual gradCPT (800 ± 30 ms) than for the auditory gradCPT ($1{,}075 \pm 30$ ms): $F(1, 22) = 42.34$, $\eta_p^2 = 0.66$, $p < .001$. The main effect of SOA was significant at 692 ± 10 ms for 800 ms SOA, 963 ± 25 ms for 1,200 ms SOA, and $1{,}157 \pm 35$ ms for 1,600 ms SOA; $F(2, 44) = 178.48$, $\eta_p^2 = 0.89$, $p < .001$. The interaction was also significant: $F(2, 44) = 8.63$, $\eta_p^2 = 0.28$, $p = .001$. The lower d' and longer RTs in the auditory gradCPT are consistent with a previous finding (Baker et al., 1995). The longer RT may be partly because the participants needed to integrate information over time, which means that they could not access the full stimulus information at the very beginning of the stimulus presentation. For the equivalence of performance level (d', approximately 3.0), we chose in Experiment 2, SOAs of 1,600 and 800 ms for the auditory and visual gradCPTs, respectively. How other related indices (hit rate, C', and RT CV) depend on SOAs is shown in Figure 3b to e.

Experiment 2: similarity of auditory and visual attentional fluctuations

We investigated auditory and visual gradCPT performance using paired *t*-tests. The results showed that d' differed between the auditory and visual gradCPT (2.85 ± 0.15 and 3.25 ± 0.14): $t = 2.75$, Cohen's $d = 0.51$, $p = .010$. As expected, however, we obtained high d' (around 3.0) from both the gradCPTs. The hit rate was better for the visual gradCPT (0.988 ± 0.003) than for the auditory CPT (0.934 ± 0.011): $t = 5.01$, Cohen's $d = 1.26$, $p < .001$. The FA rate was worse for the auditory CPT (0.236 ± 0.023) than for the visual gradCPT (0.154 ± 0.022): $t = 4.59$, Cohen's $d = 0.67$, $p < .001$. It should be noted that moderate interindividual variations of the gradCPT performance were found.

We performed a correlation analysis to investigate what behavioural indices reflect the similarity of the auditory and visual gradCPTs. We found a positive correlation between d' of the gradCPTs: $r = .51$, $p = .005$ (Figure 4a). FA rates for the auditory gradCPT were highly correlated with those for the visual gradCPT: $r = .69$, $p < .001$ (Figure 4b). However, the correlation between hit rates of the gradCPTs did not reach statistical significance: $r = .27$, $p = .16$ (Figure 4c). The correlation coefficient between FA rates in the auditory and visual gradCPTs was greater than that between the hit rates: $Z = 2.01$, $p = .043$. In addition to d' and FA rate, C' and RT coefficients of variation were also characteristic indices for the visual gradCPT (Fortenbaugh et al., 2015). Figure 4d and e shows positive correlations of C' ($r = .53$, $p = .003$) and the RT coefficient of variation ($r = .37$, $p = .050$), respectively. Our results indicate that the FA rate and its related indices (not hit rate) are important indices to bridge the gap between auditory and visual sustained attention.

To characterise the pattern of attentional fluctuations, we performed a Fourier analysis of auditory and visual VTCs. For each task, each participant's VTCs were Fourier-transformed to a frequency spectrum. Among the peak frequencies of the spectrum, the peak frequency with the largest power was defined as the VTC fluctuation frequency in each task (Figure 5a). The VTC fluctuation frequency ($M \pm SE$) was 0.020 ± 0.002 Hz for audition and 0.016 ± 0.001 Hz for vision. The visual and auditory fluctuation frequencies of participants were correlated with each other: $r = .40$, $p = .030$ (Figure 5b). The fluctuation frequency in the auditory gradCPT was similar to that in the visual gradCPT even though SOAs differed between the two gradCPTs. This suggests that the auditory and visual sustained attention shares some common principle that specifies the rhythm of each individual.

Experiment 2: relations between attention failure and RT variability

In the visual gradCPT study (Esterman et al., 2013), the rationale for using VTC as a measure of attentional fluctuation was demonstrated as the relationship between RT variability and attentional failures. We computed time-dependent changes in FA rates and RT CVs to examine the cognitive demands of gradCPTs further. Figure 6a shows

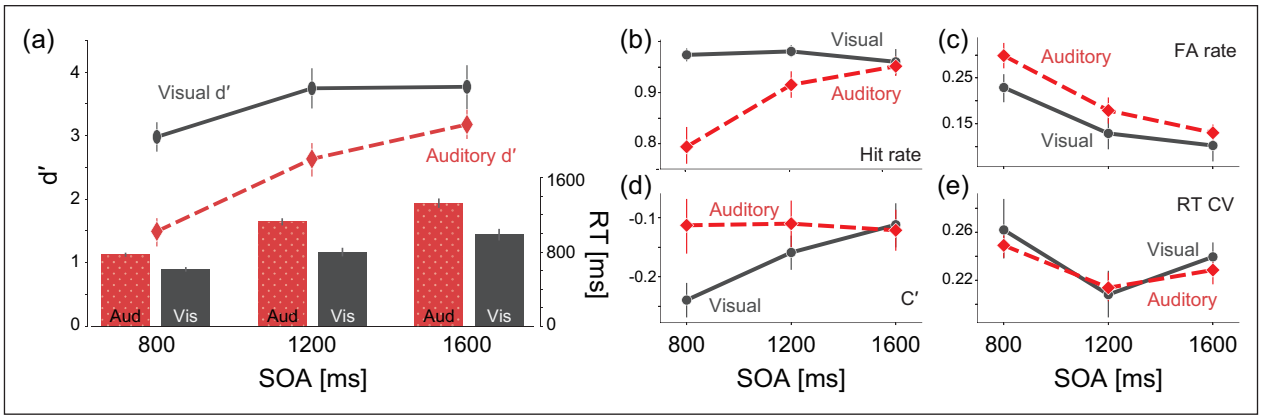

Figure 3. Effects of SOAs in Experiment 1. (a) Lines and bar charts indicate the sensitivity statistic (*d′*) and mean RTs, respectively, as a function of SOAs. (b–e) Lines indicate (b) FA rate, (c) hit rate, (d) *C′*, and (e) RT CV as functions of SOAs, respectively.
Red: auditory gradCPT. Grey: visual gradCPT. Error bars represent standard errors of means. SOA: stimulus onset asynchrony.

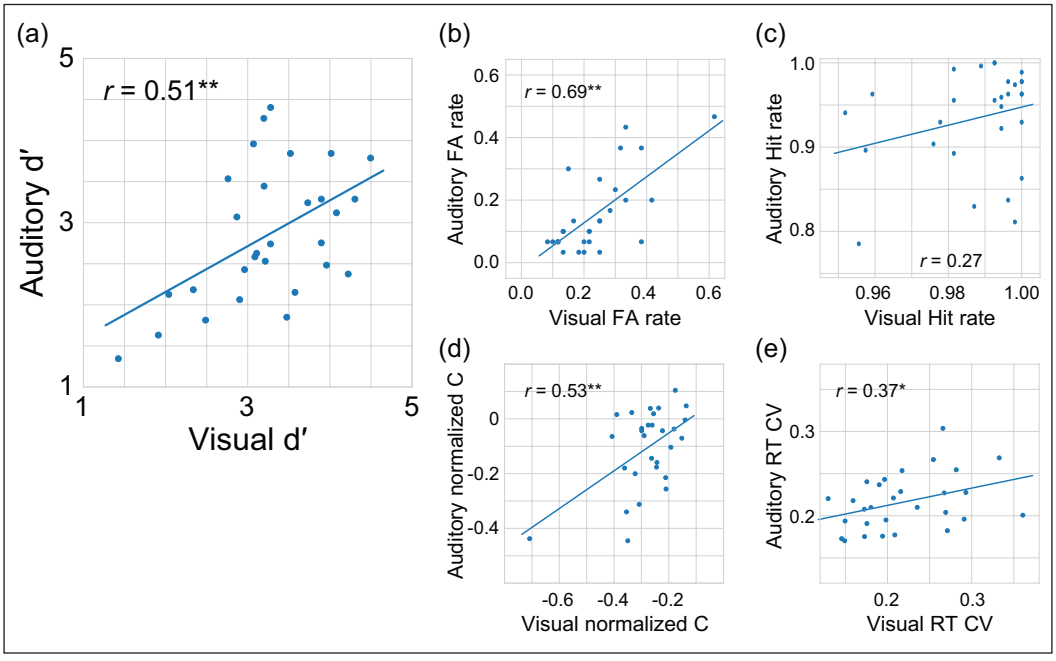

Figure 4. Correlations between auditory and visual gradCPTs in Experiment 2. Positive correlations were found for (a) *d′* and (b) false alarm (FA) rate, but not for (c) hit rate. Positive correlations were also found for (d) *C′* and (e) RT coefficient of variation. Lines show linear regressions.

the time course of FA rates and RT CVs in the sliding 2-min windows. A trend test demonstrated that FA rates and RT CVs increased over time regardless of sensory modalities. For audition and vision, FA rates increased over time ($p < .05$) and RT CVs increased over time ($p < .001$). In addition, we quantified the amount of increase as the average of the last 2 min minus that of the first 2 min. The amount of increase is also correlated between FA rates and RT CVs regardless of sensory modalities ($r = .45$, $p = .014$ for the auditory gradCPT; $r = .39$, $p = .037$ for the visual gradCPT) (Figure 6b). The results demonstrate a similarity of temporal changes between attentional failures and RT variabilities regardless of sensory modalities, which was originally reported in a vision study (Esterman et al., 2013).

We collapsed the time course of FA rates and RT CVs and confirmed the relationship between the two indices (Figure 7). FA rates were positively correlated with RT CVs: $r = .52$, $p = .004$ for the auditory gradCPT and $r = .44$, $p = .016$ for the visual gradCPT. The results indicate that the stability of RTs is linked to the reduction of attention failures, regardless of sensory modalities. Thus, both too fast and too slow RTs are considered as signatures of the lack of attention.

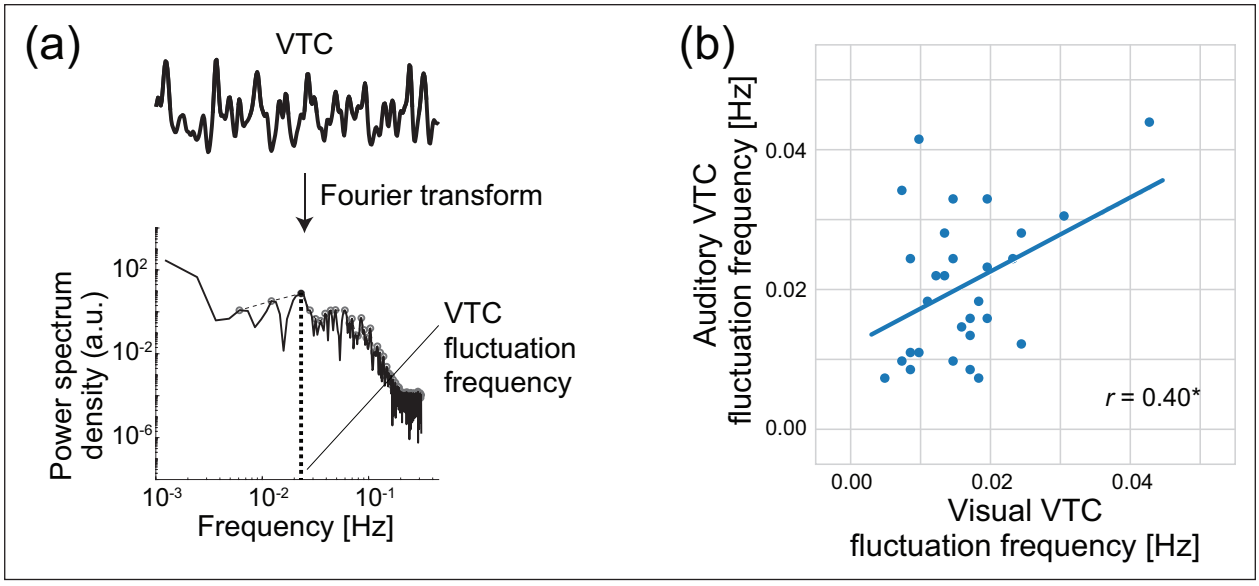

Figure 5. Fluctuation frequencies of auditory and visual gradCPTs in Experiment 2. (a) Evaluation of VTC fluctuation frequency. For each task, a Fourier analysis was used for the smoothened VTC. The peak with the largest power was defined as the VTC fluctuation frequency of the task. (b) Positive correlation of VTC fluctuation frequencies between auditory and visual gradCPTs. Smirnov–Grubbs test did not reveal any outliers. The line shows linear regression.

Discussion

The present study examined dynamic fluctuations of auditory and visual sustained attention, in addition to the temporally averaged performance, for each individual. In Experiment 1, we found changes in CPT performance with increasing SOAs. In Experiment 2, we found positive correlations of FA rates and fluctuation frequencies between the auditory and visual gradCPTs. Taken together, the results suggest that the dynamics of sustained attention in the auditory and visual modalities are underpinned by common principles.

The commonality of temporally averaged performance between the auditory and visual sustained attention is consistent with previous findings, showing a positive correlation of FA rates between auditory and visual sustained attention to response tasks (Seli et al., 2012). More importantly, we have demonstrated that the temporal dynamics of auditory and visual sustained attention are similar to each other. It should be noted that timescales of auditory and visual attentional fluctuations (from 50 to 60 s) were within a similar range, even though SOAs differed between the auditory and visual gradCPTs. Interestingly, the timescale is much longer than that of other perceptual phenomena such as multistable perception (from several to 10 s) (Kondo et al., 2018). Therefore, our results reveal new temporal aspects of auditory and visual sustained attention.

Our results support models in which the attentional fluctuation is derived from a central, modality-general system such as the resource-control model (Thomson et al., 2015). Several researchers have argued that modality-specific attentional resources exist when target detection and identification tasks are used (Alais et al., 2006; Duncan et al., 1997; Larsen et al., 2003). However, these tasks lack processes to continuously update target information. In other words, the target detection/identification tasks may have low attentional demands because participants only pay attention to a certain stimulus dimension (Duncan & Humphreys, 1989). Thus, there is the possibility that detection and identification performance does not reflect the temporal dynamics of attentional control systems, but does reflect time-invariant features (Fougnie et al., 2018). In contrast, our gradCPTs require high attentional demands throughout a whole sequence of trials. A previous study using multiple object tracking tasks showed that auditory and visual tracking shares attentional resources (Fougnie et al., 2018). Therefore, the temporal dynamics of sustained attention are probably governed by principles not specific to a certain modality.

Our analyses on the time-on-task effects (Figure 6) also support the resource-control model (Thomson et al., 2015). The model predicts that less attentional resources are allocated in later trials because the total amount of resources is fixed and the controller allocates more resources into the default mind-wandering. In accordance with that prediction, our result showed similar trends in terms of the FA rate and RT CV (Figure 6a). The concept of a higher level system like the controller can be discussed with other theories of temporal aspects of attention such as dynamic attending (Jones, 2019), where voluntary manipulation of a driving rhythm is essential

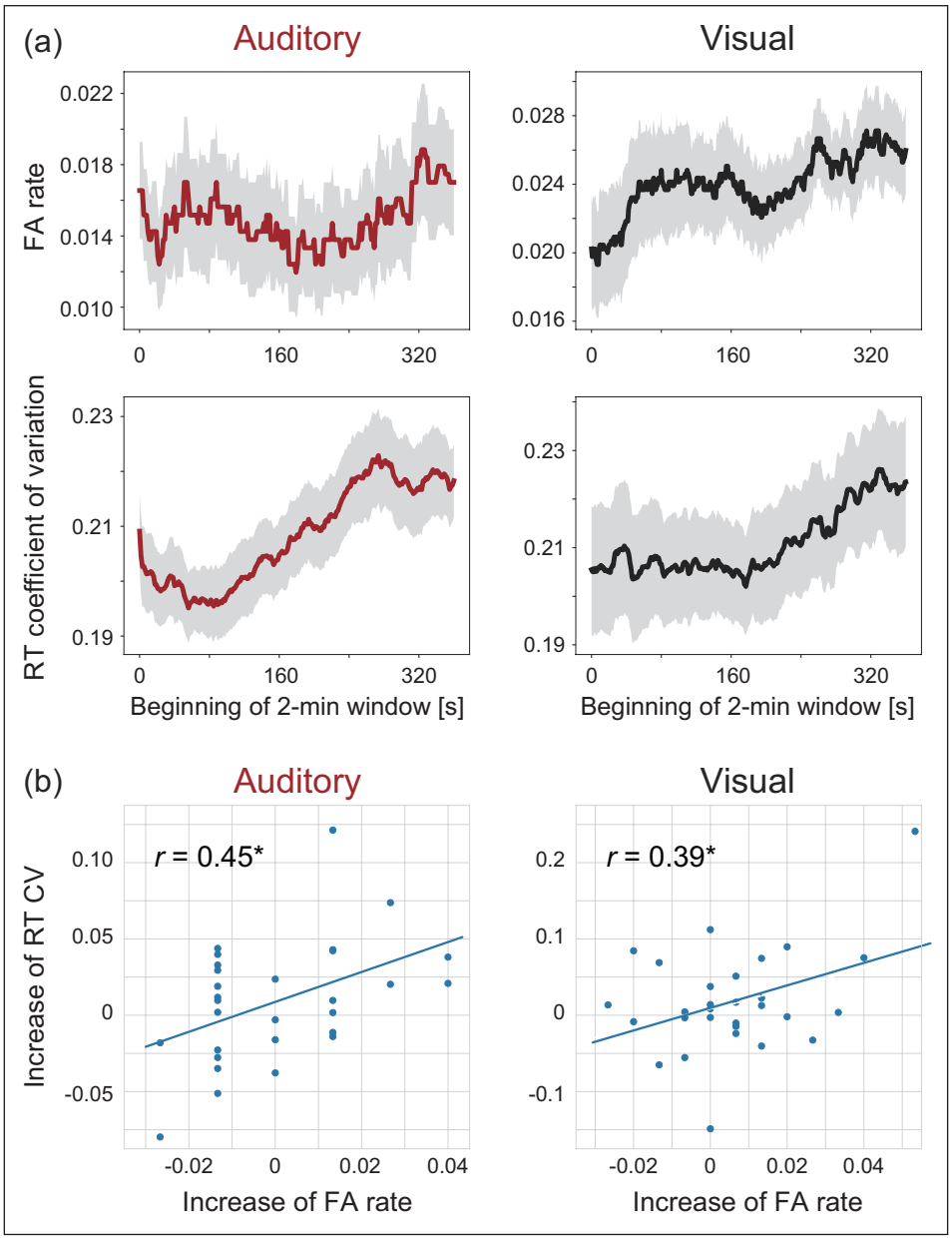

Figure 6. Temporal changes in the task accuracy and speed stability. (a) The FA rate and the RT coefficient of variation increase gradually over time, regardless of task modalities (trend tests, $p < .05$ in all the cases). Grey areas indicate the standard errors of means. (b) The amount of increase (the last 2 min minus the first 2 min) correlates between the FA rate and the RT coefficient of variation, regardless of task modalities ($p = .014$ for the auditory gradCPT; $p = .037$ for the visual gradCPT).

and probably modality-general. Besides, the within-modality link between the FA rate and RT CV (Figure 6b) supports our gradCPT as an auditory analogue of the visual gradCPT. Recent attention studies have started to discuss the fluctuation of motor outputs even without any sensory stimuli (Kucyi et al., 2017) or the interaction of sensory and motor modalities (Zalta et al., 2020), which could enable us to examine the extent to which the controller is modality-general. Our auditory gradCPT broadens ways of between-modality comparison in terms of the attentional dynamics.

Another model could account for the attentional dynamics from a different perspective. Specifically, the opportunity cost model (Kurzban et al., 2013) argues that choosing to sustain the attention to a particular task means not choosing the next-best task (in our case, mind-wandering). This choice is explained by comparing the costs (especially the opportunity cost of not choosing the next-best task) with the benefits. The overall performance decrement arises from the decreasing relative utility of the imposed task. Following this model, the modality-general aspect of sustained attention could be explained by a

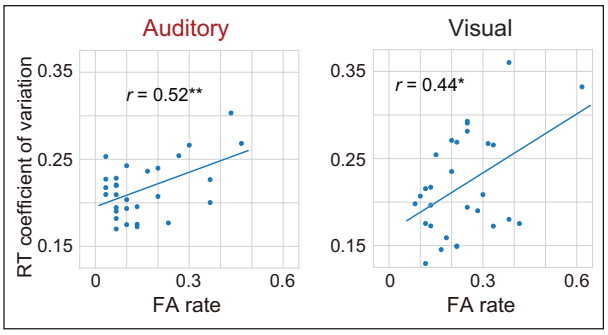

Figure 7. Individual differences in the relationship between the FA rate and the stability of RTs. Correlations between the FA rate and RT coefficient of variation are significant ($p < .05$). Circles indicate individual data points in the auditory and visual gradCPTs.

modality-general strategy for the utility comparison. The attentional fluctuations we focused on in this study might also be explained by fluctuations in the estimated relative utility. Although it is difficult to formulate the utility specifically, the opportunity-cost model provides a suggestion for the understanding of the attentional dynamics.

Our results suggest some common principles underlying auditory and visual sustained attention, but not the existence of a single neural locus that controls sustained attention in both modalities. Neural processing for auditory sustained attention may be shared with that of the visual gradCPT (Rosenberg et al., 2017). However, it is also possible that each has distinct neural circuits that are still underpinned by common principles. This reminds us that auditory and visual perceptual organisation is implemented independently across modalities but modulated by similar principles of neural competition (Kondo et al., 2012; Pressnitzer & Hupé, 2006). A previous study reported that there is some overlap in the attentional control of auditory and visual modalities (Talsma et al., 2008), although it did not investigate dynamic fluctuations. Neural underpinnings of the dynamic aspects have been mainly studied using the visual gradCPT (Esterman et al., 2013; Fortenbaugh et al., 2018). Combined with another paradigm without sensory stimuli (Kucyi et al., 2017), the results have shown that the dorsal attention network and the default mode network are the networks involved in modality-general sustained attention. Additional tests for the modality-general aspects will be enabled by the auditory gradCPT.

A concern might be that some of the similarities we demonstrated could be affected mainly by factors other than sustained attention, for instance, general factors such as cognitive ability or the equalisation of difficulty in Experiment 1. Indeed, our experiments cannot exclude this possibility completely, but we think that this interpretation is unlikely for the following reasons. First, in Experiment 1, the fluctuation frequencies did not differ between SOAs of 800 and 1,600 ms (two-sided paired t-test: $t = 1.69$, $p = .12$ for the auditory gradCPT; $t = 0.83$, $p = .42$ for the visual gradCPT). Thus, the similarity of the fluctuation frequency was not due to the difficulty equalisation. Next, the RT coefficient of variation, another measure to assess the fluctuation, was not equalised by the difficulty equalisation (Figure 3e). This indicates that the difficulty equalisation did not eliminate all attentional differences. Finally, the difficulty equalisation was done based on the average performance of two distinct populations. This does not necessarily lead to correlations at the level of individuals we showed in Figures 4 and 5. Taken collectively, our results suggest that the similarities are diagnostic of common underlying mechanisms of sustained attention, although we do not exclude the possible contributions of top-down effects like the general factors or the bottom-up biological rhythm from the brainstem (Kondo et al., 2012).

To enable direct comparisons with the visual gradCPT (Esterman et al., 2013), in the present study we followed their procedure as far as we could. It includes defining the RT and smoothing the VTCs, all of which depend on an arbitrary boundary or threshold values (e.g., 70% and 40% for RTs). Slight changes in the values did not affect the main arguments in the present study.

The present study established a new platform for investigating dynamic aspects of resource allocation in sustained attention. We used auditory and visual gradCPTs with relatively simple stimuli. Our approach makes it possible to examine how the two modalities interact in a dual-task paradigm or how fluctuation patterns can be changed by stimuli associated with specific types of values. Some previous studies explored neural correlates of sustained attention (Esterman et al., 2013; Rosenberg et al., 2016), but the general interpretation was limited because the modality of the tasks used was typically only vision. Cross-modal investigations on vulnerable sustained attention could contribute to a better understanding of our adaptive behaviours.

Acknowledgements

The authors thank the editor and two anonymous reviewers for their thoughtful comments on an earlier version of this manuscript. They also thank the Institute for Advanced Collaborative Research at Chukyo University for its generous support.

Author contributions

All authors conceived of and designed the study. J.I.K. and H.M.K. conducted the experiments. H.T., K.K., and H.M.K. analysed the data. All authors interpreted the results and wrote the manuscript.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by JSPS KAKENHI Grants (Nos 17K04494 and 20H01789 to K.K., J.I.K., and H.M.K.).

ORCID iDs

Hiroki Terashima https://orcid.org/0000-0002-8825-5299
Hirohito M. Kondo https://orcid.org/0000-0002-7444-4996

References

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1339–1345. https://doi.org/10.1098/rspb.2005.3420

Aylward, G. P., Brager, P., & Harper, D. C. (2002). Relations between visual and auditory continuous performance tests in a clinical population: A descriptive study. *Developmental Neuropsychology*, *21*, 285–303. https://doi.org/10.1207/S15326942DN2103_5

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baker, D. B., Taylor, C. J., & Leyva, C. (1995). Continuous performance tests: A comparison of modalities. *Journal of Clinical Psychology*, *51*, 548–551. https://doi.org/10.1002/1097-4679(199507)51:43.0.CO;2-Q.

Ballard, J. C. (2001). Assessing attention: Comparison of response-inhibition and traditional continuous performance tests. *Journal of Clinical and Experimental Neuropsychology*, *23*, 331–350. https://doi.org/10.1076/jcen.23.3.331.1188

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446. https://doi.org/10.1163/156856897X00357

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. https://doi.org/10.1037/0033-295x.96.3.433

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810. https://doi.org/10.1038/42947

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*, 2712–2723. https://doi.org/10.1093/cercor/bhs261

Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, *1396*, 70–91. https://doi.org/10.1111/nyas.13318

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*, *26*, 1497–1510. https://doi.org/10.1177/0956797615594896

Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J., & Esterman, M. (2018). Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension. *NeuroImage*, *171*, 148–164. https://doi.org/10.1016/j.neuroimage.2018.01.002

Fougnie, D., Cockhren, J., & Marois, R. (2018). A common source of attention for auditory and visual tracking. *Attention, Perception, & Psychophysics*, *80*, 1571–1583. https://doi.org/10.3758/s13414-018-1524-9

Galinsky, T. L., Rosa, R. R., Warm, J. S., & Dember, W. N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors*, *35*, 603–614. https://doi.org/10.1177/001872089303500402

Imbert, J. P., Hodgetts, H. M., Parise, R., Vachon, F., Dehais, F., & Tremblay, S. (2014). Attentional costs and failures in air traffic control notifications. *Ergonomics*, *57*, 1817–1832. https://doi.org/10.1080/00140139.2014.952680

International Phonetic Association. (1999). *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*. Cambridge University Press.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review*, *6*, 87–92. https://doi.org/10.3758/BF03210813

Jones, M. R. (2019). *Time will tell: A theory of dynamic attending*. Oxford University Press.

Kim, S., Park, Y., & Headrick, L. (2018). Daily micro-breaks and job performance: General work engagement as a cross-level moderator. *Journal of Applied Psychology*, *103*, 772–786. https://doi.org/10.1037/apl0000308

Kondo, H. M., Kitagawa, N., Kitamura, M. S., Koizumi, A., Nomura, M., & Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cerebral Cortex*, *22*, 1915–1922. https://doi.org/10.1093/cercor/bhr266

Kondo, H. M., Pressnitzer, D., Shimada, Y., Kochiyama, T., & Kashino, M. (2018). Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. *Scientific Reports*, *8*, 14548. https://doi.org/10.1038/s41598-018-32892-3

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral Cortex*, *27*, 1831–1840. https://doi.org/10.1093/cercor/bhw029

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*, 661–679. https://doi.org/10.1017/s0140525x12003196

Larsen, A., McIlhagga, W., Baert, J., & Bundesen, C. (2003). Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. *Perception and Psychophysics*, *65*, 568–574. https://doi.org/10.3758/bf03194583

Lehnert, G., & Zimmer, H. D. (2008). Common coding of auditory and visual spatial information in working memory. *Brain Research*, *1230*, 158–167. https://doi.org/10.1016/j.brainres.2008.07.005

MacLean, K. A., Aichele, S. R., Bridwell, D. A., Mangun, G. R., Wojciulik, E., & Saron, C. D. (2009). Interactions between endogenous and exogenous attention during vigilance. *Attention, Perception, & Psychophysics*, *71*, 1042–1058. https://doi.org/10.3758/APP.71.5.1042

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion*, *31*, 5–11. https://doi.org/10.1007/s11031-007-9053-1

Moore, T., & Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual Review of Psychology*, *68*, 47–72. https://doi.org/10.1146/annurev-psych-122414-033400

Nobre, A. C. (2018). Attention. In J. T. Wixted & J. Serences (Eds.), *Sensation, perception, and attention: Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., *Vol. 2*, pp. 241–315). Wiley.

Parasuraman, R., & Giambra, L. (1991). Skill development in vigilance: Effects of event rate and age. *Psychology and Aging*, *6*, 155–169. https://doi.org/10.1037/0882-7974.6.2.155

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366

Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, *16*, 1351–1357. https://doi.org/10.1016/j.cub.2006.05.054

Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing attention with predictive network models. *Trends in Cognitive Sciences*, *21*, 290–302. https://doi.org/10.1016/j.tics.2017.01.011

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*, 165–171. https://doi.org/10.1038/nn.4179

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, *136*, 663–684. https://doi.org/10.1037/0096-3445.136.4.663

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. *Canadian Journal of Experimental Psychology*, *66*, 44–50. https://doi.org/10.1037/a0025111

Sturm, W., & Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. *NeuroImage, 14*, S76–S84. https://doi.org/10.1006/nimg.2001.0839

Szalma, J. L., Warm, J. S., Matthews, G., Dember, W. N., Weiler, E. M., Meier, A., & Eggemeier, F. T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors*, *46*, 219–233. https://doi.org/10.1518/hfes.46.2.219.37334

Talsma, D., Kok, A., Slagter, H. A., & Cipriani, G. (2008). Attentional orienting across the sensory modalities. *Brain and Cognition*, *66*, 1–10. https://doi.org/10.1016/j.bandc.2007.04.005

Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science*, *10*, 82–96. https://doi.org/10.1177/1745691614556681

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*, 433–441. https://doi.org/10.1518/001872008X312152

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*, 449–455. https://doi.org/10.1518/001872008X288394

Zalta, A., Petkoski, S., & Morillon, B. (2020). Natural rhythms of periodic temporal attention. *Nature Communications*, *11*, 1051. https://doi.org/10.1038/s41467-020-14888-8