5-1-2021

Exploring the Relation Between Distractor Inhibition and Audiovisual Integration

Michelle H. Wang
Wright State University - Main Campus, wang.202@wright.edu

Elizabeth L. Fox

EXPLORING THE RELATION BETWEEN DISTRACTOR INHIBITION AND AUDIOVISUAL INTEGRATION

Michelle H. Wang, Wright State University, Dayton, OH

Elizabeth L. Fox, Air Force Research Laboratory, WPAFB, OH

The Stroop paradigm is a great experimental tool to assess the extent that task-irrelevant, but target-related, distractors influence target identification in a variety of contexts. In particular, it has been applied beyond the traditional visual modality (e.g., audio, or audiovisual). However, audiovisual studies using Stroop-like tasks have reported conflicting results. Importantly, these bimodal studies assessed only group-level mean differences and did not investigate whether the degree of bimodal conflict is greater than what is expected of two unimodality distractors that are inhibited in an unlimited capacity, independent, and parallel fashion. In this research, we relied on cognitive-based models of individuals' performance to estimate audiovisual conflict and directly compared the influence that two types of bimodal distractors had on performance: 1) the same conflicting information was presented in both modalities and 2) different conflicting information was presented in each modality. We found unimodal visual, but not auditory, distractors significantly influenced target processing. Most interestingly, we found that despite a lack of unimodal auditory influence some participants performance indicated that bimodal distractors were harder (easier) to inhibit than expected given our model-based predictions, and the direction (limited or super capacity) and degree of deviation from our model prediction depended on cross-modal distractor similarity.

The Stroop Effect (Stroop, 1935) is a phenomenon demonstrating the compelling interference of semantics (i.e., color words) on the identification of the font color of the same word. It also provides insight into one's ability to inhibit such interferences. More recently, interest in the Stroop Effect has expanded into the auditory modality. However, there has been conflicting results in regard to auditory interference: some found evidence of auditory interference in addition to visual interference (Cowan & Baron, 1987), others observed evidence that auditory distractor did not provide additional interference beyond that of visual distractors (Elliott et al., 2014). Perhaps other factors could modulate auditory interference in a Stroop-like paradigm. Francis, McLeod, and Taylor (2014) specifically examined the relationship between interference and distractor similarity: distractors in different modalities both provided color information incongruent to the font color, but the information was either the same across modalities or different. They compared the effect of incongruent distractors to that of the control condition with only uninformative stimuli (i.e., visual control: "xxxx", auditory control: tone). Interference was calculated by subtracting the mean reaction time (RT) of the control condition from the mean RT of each of the incongruent conditions and averaged at the group-level. They found that the presence of a distractor, whether auditory or visual, increased interference as compared to the control regardless of whether the distractors from different modalities were the same or different (agrees with Cowan & Baron, 1987, but disagrees with Elliott et al., 2014). Furthermore, Francis et al. observed that the combined interference in the incongruent same distractors condition were less than that of the sum of the unimodal interferences, whereas the combined interference in the incongruent different condition was approximately equal to that of the sum of the unimodal interferences. This suggests that distractors from both modalities may have integrated when they were the same but may have been processed independently when they were different.

In the currently study, we leverage one measure, the capacity coefficient, from an established mathematical modelling framework, system factorial technology (SFT; see Townsend & Nozawa, 1995 for details), to assess processing capacity to inhibit distractors in a font color judgment where distractors vary in number and modality, at the individual-level. Theoretically, the color judgment in the Stroop task can be made without processing any distractors. If this were the case, using the cumulative reverse hazard function, *K*, at any time, *t*, we would expect that processing times to make a color judgment, *C*, would be the same regardless of the distractors, *i*. However, the Stroop effect shows that the processing time of the font color may speed-up (when written word semantic and font color are congruent) or slow-down (when semantics and font color are incongruent). This discrepancy can be computed as a capacity measure from SFT called a single-target self-terminating process (ST-ST; Blaha, 2010): $C(t)_{STST} = \frac{K_C}{K_{C(i)}}$. A result equal to 1 indicates no change in color judgement processing capacity (*unlimited capacity*). Otherwise, processing times would speed up (greater than 1; *super capacity*) or slow down (less than 1; *limited capacity*) depending on the distractor information. Analogously, the same ST-ST equation can assess color judgement processing time with auditory distractors by substituting visual distractors with non-informative characters and introduce auditory distractors.

Distractors may also be bimodal, that is, spoken and written color words incongruent with the font color. Therefore, we were also interested in modeling the efficiency to ignore bimodal distractors, and the potential effect of bimodal distractor similarity. One's efficiency to inhibit distractors can be defined as the residual cost of inhibition for each unimodal condition by subtracting the processing time to make the color judgement alone from the total processing time. Using the capacity-AND decision-rule (i.e., both distractors must be inhibited to make a decision; Townsend, 1974), we formed a model prediction of bimodal performance that assumes distractors are processed with Unlimited Capacity, and in an Independent and Parallel fashion (UCIP). The capacity-AND measure is the ratio of the cumulative reverse hazard function of response times of bimodal distractors at a given time, *t*, for the color of the word, *C*, the written visual word, *V*, and the auditory spoken word, *A*. We do not assume that the written and spoken word need to be processed but define the degree to which their processing occurs depending on the degree to which response times slowed in the single-modality distractor conditions, $K_{C(V)}$ and $K_{C(A)}$. The UCIP model baseline predicts the sum of the processing time to allow the color to influence their judgment, K_C, and the processing time to inhibit written or spoken word interference, $K_V + K_A$, should equal the cumulative reverse hazard function for the combination of the two distractors, $K_{C(AV)}$. The cost of processing time to inhibit distractors was not directly observable. However, we could estimate it by accounting for the processing time of making the color judgment alone with no distractors, compared to in-context of the visual written word to obtain $K_V(t)$. Likewise, for the auditory spoken word $K_A(t)$. Therefore, we can obtain our UCIP prediction to compare to observed performance using a capacity-AND form:

$$C_{AND} = \frac{K_{C(V)}(t) + K_{C(A)}(t) - K_C(t)}{K_{C(AV)}}.$$

Like the ST-ST measure, performance is characterized as *limited* (slower to inhibit bimodal distractors than UCIP model predictions), *unlimited* (as predicted), or *super capacity* (faster).

The current study provided several new contributions to the methodology of the existing literature. We examined the effect of unimodal and bimodal distractors: 1) across the entire distribution of RTs, 2) at an individual-level, 3) using cognitive-model based comparisons, and 4) with approximately 10 times the amount of data compared to previous studies. We predicted a main effect of interference (i.e., distractors would slow RTs), and an interaction between the degree of interference and modality (i.e., visual > auditory). We also predicted performance with bimodal distractors would deviate from UCIP model predictions, and the degree of violation would depend on whether the spoken and written color words were the same or different.

Methods

This study was administered virtually and completed at the subjects' times and locations of choice using the Amazon WorkSpaces, which is a virtual desktop that requires internet connection (experimenters were available virtually). The testing environment and equipment were kept consistent within each subject: the task was generated and administered via PsychoPy3 (Pierce, 2007), the auditory stimuli were delivered via subject's headphones of choice, and subjects used a keyboard to respond. Twelve subjects (reported normal hearing, normal or corrected-to-normal vision, and normal color vision) participated in this experiment. Ten long-term subject panel members (Wright-Patterson Air Force Base) were compensated at an hourly rate, two recruits were compensated at $15/hour, and one was an author of this paper.

Table 1.

Audiovisual Stroop Trial Types

Condition	Stimulus Example		
	Auditory	Visual	Font Color
Control	**<noise>**	**@@@@**	**Blue**
		@@@@	Blue
Congruent			
Visual		blue	Blue
Visual	<noise>	blue	Blue
Auditory	blue	@@@@	Blue
VA	blue	blue	Blue
Incongruent			
Visual		red	Blue
Visual	**<noise>**	**red**	**Blue**
Auditory	**red**	**@@@@**	**Blue**
VA Same	**red**	**red**	**Blue**
VA Different	**red**	**green**	**Blue**
Mixed			
Visual Incongruent	blue	red	Blue
Audio Incongruent	red	blue	Blue

Note. Conditions of interest are bolded.

The visual stimuli included words "red", "green", "blue", and the symbol set "@@@@", which were presented for 650 ms in Arial font, with either red, green, or blue font color at 0.2 normalized letter height (scaled with monitor settings). The auditory stimuli were spoken color words from the same female speaker (i.e., "red", "green", "blue") and white noise delivered at a comfortable listening level. All auditory stimuli were 650 ms in duration, except for "red", which was 601 ms. Subjects were instructed to set the system sound level to the highest setting but were able to adjust the sound settings at their own will. This study was a part of a larger project, which included all trials types shown in Table 1. However, only five trial types (bolded in Table 1) were analyzed in the current study: control, visual, auditory, incongruent same (IS), and incongruent different (ID). The objective was to correctly identify the font color (target) while written and/or spoken words (distractors) were present. Each trial was 4000 ms with the same order of events: a visual fixation cross and an auditory fixation tone (500 ms), randomized inter-stimulus interval (250-500 ms), observation interval (650 ms), response interval (2000-2250 ms; varied duration does not affect response, $RT_{90th\ percentile}$ = 1048 ms), inter-trial interval (1000 ms). Subjects each completed 160 blocks (randomized conditions) over 10 experimental session. All experimental session included conditions from the larger project (using the same set of stimuli as the current study) the data of which is not discussed in this paper and should not have altered one's performance in the trials of interests. Fifteen-second breaks were enforced between blocks, but subjects were able to take longer breaks if needed as each block was self-initiated by a key press any time after the enforced break. Prior to each session, subjects completed a 72-trial practice, which included all possible trials present in the experimental session.

Results and Discussion

Performance was highly accurate, 96.7% of the trials were correct. Only correct responses, greater than 100ms in duration, were kept for further analysis. All but one subject showed significantly slowed capacity to process color information in context of visual distractors (written word), $C_{Z\,(V)}$ = [-9.56, -1.50], M = -3.71, successfully replicating the traditional (visual) Stroop effect. But only one subject showed a significant change from unlimited capacity with auditory distractors (spoken words), $C_{Z\,(A)}$ = [-2.24, 1.68], M = -0.37. In general, subjects could more easily inhibit auditory distractor as compared to visual distractor carrying the same information. Some cognitive control seemed necessary to inhibit visual distraction (evidenced by slower RT), which may result from limited capacity processes, an inefficient system structure (i.e., serial processing of each piece of information) or interdependent pooling (i.e., coactive architecture) of the target and visual distractor information (Little, Eidels, Fifić, & Wang, 2018). We found unlimited capacity processing to inhibit auditory distractors, which suggests efficient inhibition of spoken words.

Figure 1 shows individualized estimates of the processing capacity to inhibit both auditory and visual distractors relative to the UCIP model prediction (black solid line, where $C(t) = 1$). In general, the subjects' processing capacity fall into three groups: unlimited, limited, and similarity dependent. For some (*Fig*. 1a), we found unlimited processing capacity for both distractor similarity types: incongruent-same (IS), $C_{Z\,(\mathrm{IS})}$ = [-1.03, 1.32], $M_{(\mathrm{IS})}$ = -0.16, and incongruent-different (ID), $C_{Z\,(\mathrm{ID})}$ = [-1.54, 0.45], $M_{(\mathrm{ID})}$ = -0.72, which follows from an unlimited capacity, independent, and parallel processing structure. These findings are explained with principles of multisensory integration: multisensory enhancement more often occurs when stimuli from different modalities are presented from the same spatial location, within the same time interval, and/or has similar effectiveness (Holmes & Spences, 2005; Meredith & Stein, 1983). There are several factors in the current study that could have violated these principles and hindered integration: 1) the visual (monitor) and the auditory (headphones) stimuli were not spatially co-located, 2) semantic processing of the distractors were not temporal aligned: the written word was presented instantaneously whereas it takes time to deliver the entire spoken word, and 3) perhaps most importantly, there was a clear modality asymmetry in distractor inhibition: the visual distractor was far more effective (i.e., more difficult to inhibit) than the auditory distractor.

Another group of subjects (*Fig*. 1b) showed significantly slowed processing capacity than the UCIP model prediction (limited) in both IS and ID conditions ($C_{Z\,(\mathrm{IS})}$ = [-3.31, -2.28], $M_{(\mathrm{IS})}$ = -

2.70; $C_{Z\,(ID)}$ = [-2.84, -2.21], $M_{(ID)}$ = -2.50). For this group, the presence of two distractors slowed down processing regardless of whether the distractors were the same or different from one another. Here, limited capacity performance may result from processes that depend on distractors similarity: written and spoken words that were different from one another may result from serial processes (one at a time) and hence slow response times. Alternatively, identical written and spoken distractors may combine to make a stronger composite distractor due to integration or a coactive processing architecture and result in limited capacity performance.

A final group of subjects exhibited processing capacity depended on distractor similarity (similarity-dependent; *Fig*. 1c). Subject 9007 in this group showed unlimited capacity with ID trials ($C_{Z\,(ID)}$ = -1.51) and limited capacity in IS conditions ($C_{Z\,(IS)}$ = -2.15), suggesting potential bimodal integration when the distractors shared the same information. Alternatively, Subject 9001 exhibited the opposite pattern, $C_{Z\,(ID)}$ = -2.49 (limited), $C_{Z\,(IS)}$ = -1.30 (unlimited).

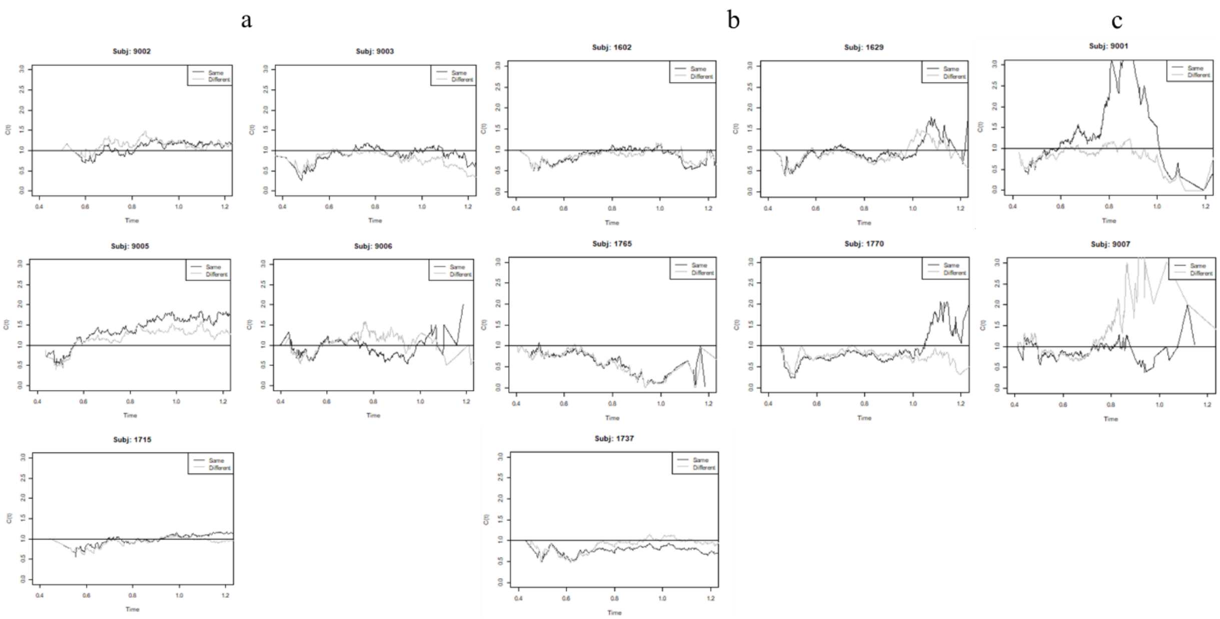

Figure 1. Individual estimates of processing capacity to inhibit same/different bimodal distractors.

Our statistical test assessed capacity across the full function. Further visual inspection of the functions indicated a potential shift in processing capacity between early and late processing times (around 800ms) for some subjects. Specifically, capacity for inhibiting distractors at later processing times increased (i.e., super capacity). Audiovisual integration may have occurred only during later processing times for these subjects when the auditory distractor was fully processed, integrated into a composite and bimodal distractor, and more efficiently inhibited. As discussed previously, the spoken words (and the semantic meaning) took time to convey, but written words were presented instantaneously. We will investigate this in a future study by shifting forward the spoken word onset and by conducting separate statistical tests for early and late processing times.

In conclusion, the current study examined audiovisual distractor inhibition within a Stroop-like paradigm and its interaction with the semantic similarity between two distractors from different modalities. We created a new measure of capacity to examine changes in processing capacity to inhibit bimodal distractors at an individual level. We observed asymmetry in the effect of unimodal distractors, specifically, it was more difficult to inhibit visual distractors than auditory distractors. Also, we found individuals' processing capacity to inhibit bimodal distractors were either: all unlimited capacity, all limited capacity, or similarity dependent. We plan to conduct follow-up studies using another measure of SFT, the survivor interaction contrast, to investigate the processing architecture of the distractors using a factorial manipulation to the processing speed of each distractor type. We will also change the temporal alignment of distractors and present the auditory distractor before the visual distractor to shorten the gap in processing time and facilitate more influence from the auditory distractor on processing times. Indeed, pilot data show incongruent spoken words presented 250ms before the onset of the target significantly slow response times in the color judgment task.

Acknowledgement

This study was supported by the Air Force Research Laboratory and Oakridge Institute of Education and Science. We would like to thank Dr. Brian Simpson for his guidance and support.

Reference

Blaha, L. M. (2011). *A dynamic Hebbian-style model of configural learning* (Doctoral dissertation, Indiana University).

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics, 41*, 393–401.

Elliott, E. M., Morey, C. C., Morey, R. D., Eaves, S. D., Shelton, J. T., & Lutfi-Proctor, D. A. (2014). The role of modality: Auditory and visual distractors in Stroop interference. *Journal of Cognitive Psychology, 26*, 15–26.

Francis, W. S., MacLeod, C. M., & Taylor, R. S. (2017). Joint influence of visual and auditory words in the Stroop task. *Attention, Perception, & Psychophysics, 79*(1), 200-211.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Current Biology, 15*,762-764.

Little, D. R., Eidels, A., Fifić, M., & Wang, T. S. (2018). How do information processing systems deal with conflicting information? Differential predictions for serial, parallel, and coactive models. *Computational Brain & Behavior, 1*, 1-21.

Meredith, M, A, & Stein, B. E. (1983) Interactions among converging sensory inputs in the superior colliculus. *Science, 221*, 389–391

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In *B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition* (p. 133-168). Hillsdale, NJ: Erlbaum Press.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 321–359.