

RESEARCH ARTICLE

When two is worse than one: The deleterious impact of multisensory stimulation on response inhibition

Kuzma Strelnikov^{1,2}*, Mario Hervault¹, Lidwine Laurent¹, Pascal Barone¹

1 Brain & Cognition Research Center (CerCo), University of Toulouse 3 –CNRS, Toulouse, France, **2** Purpan University Hospital, Toulouse, France

* These authors contributed equally to this work.

* kuzma.strelnikov@cnrs.fr



OPEN ACCESS

Citation: Strelnikov K, Hervault M, Laurent L, Barone P (2021) When two is worse than one: The deleterious impact of multisensory stimulation on response inhibition. PLoS ONE 16(5): e0251739. <https://doi.org/10.1371/journal.pone.0251739>

Editor: Katsumi Watanabe, Waseda University, JAPAN

Received: November 23, 2020

Accepted: May 1, 2021

Published: May 20, 2021

Copyright: © 2021 Strelnikov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data are in a public repository. <https://osf.io/v2b7z/>.

Funding: This study was supported by PhD funding to M. Hervault (University of Toulouse 3), recurrent CNRS funding, ANR HearCog (ANR-17-CE28-0006-01) and ANR AgeHear (ANR-21-AAP-CE28). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Multisensory facilitation is known to improve the perceptual performances and reaction times of participants in a wide range of tasks, from detection and discrimination to memorization. We asked whether a multimodal signal can similarly improve action inhibition using the stop–signal paradigm. Indeed, consistent with a crossmodal redundant signal effect that relies on multisensory neuronal integration, the threshold for initiating behavioral responses is known for being reached faster with multisensory stimuli. To evaluate whether this phenomenon also occurs for inhibition, we compared stop signals in unimodal (human faces or voices) versus audiovisual modalities in natural or degraded conditions. In contrast to the expected multisensory facilitation, we observed poorer inhibition efficiency in the audiovisual modality compared with the visual and auditory modalities. This result was corroborated by both response probabilities and stop–signal reaction times. The visual modality (faces) was the most effective. This is the first demonstration of an audiovisual impairment in the domain of perception and action. It suggests that when individuals are engaged in a high–level decisional conflict, bimodal stimulation is not processed as a simple multisensory object improving the performance but is perceived as concurrent visual and auditory information. This absence of unity increases task demand and thus impairs the ability to revise the response.

Introduction

Auditory perception and visual perception have been studied as separate channels, even though the ecological environment is rarely unimodal, and auditory and visual stimuli often combine to provide complementary information about a single multimodal object. There is now a large body of literature clearly demonstrating that integrating sensory information from different channels leads, at both perceptual and neuronal levels, to a different kind of processing from that for a single modality. At the behavioral and perceptual levels, *multisensory integration* reduces perceptual ambiguity, lowers sensory thresholds, and enhances the speed and accuracy of stimulus detection [1, 2]. The perceptual benefit of multisensory integration is

supported by direct functional interactions between sensory areas of different modalities, from the early stages of sensory processing to more integrative levels [3]. Across species, behavioral facilitation by multisensory stimuli manifests itself in shorter reaction times (RTs), compared with unimodal RTs in what is known as the *redundant signals effect* [4]. RTs for audiovisual stimulation often violate race models between sensory modalities, which predicts that the faster of the two stimuli in competition mediates the behavioral response in any given trial [5–7]. According to the co-activation model [8], which relies on multisensory neuronal integration, the threshold for initiating behavioral responses is reached faster with multisensory stimuli. The fact that this multisensory gain exceeds the facilitation predicted by summing separate response probabilities points to a form of neuronal integration, rather than simply an accumulation of sensory information [9].

Research on audiovisual integration has clearly revealed the greater efficiency of congruent audiovisual stimulation relative to unimodal stimuli, in terms of both RTs and response probabilities [10]. Most of these studies were designed to elicit a response action, such as pressing a response button in a detection, discrimination, or memorization task [11]. However, there is often a need to stop action in response to a certain signal in ecological and social situations. One quite standard approach to testing the capacity of a signal to inhibit action is the *stop-signal paradigm* [12], in which participants need to respond quickly to a specific main stimulus, but sometimes that main stimulus is followed by a stop signal, upon which participants are supposed to cancel their response. According to the horse-race model [12, 13], successful inhibition in stop-signal tasks relies on the outcome of a race between independent go and stop processes. Inhibitory control fails when the go process finishes the race before the stop one and succeeds when the stop process reaches the response threshold first. The statistical model underlying the horse-race model provides a means of estimating the duration of the covert action inhibition process, in the form of a stop-signal reaction time (SSRT). In this paradigm, inhibitory control can thus be described by both the inhibition function (i.e., probability of responding as a function of the time that elapses between the main stimulus and the stop signal) and the SSRT [14]. The ability to inhibit action is a central executive function that is crucial in various situations and plays a role in several psychopathologies [15].

However, the importance of sensory processes in inhibitory mechanisms has only recently attracted the attention of researchers. Several differences in efficiency between auditory and visual stop-signals have been reported, albeit with non-ecological stimuli [16]. In addition, authors recently found that manipulating the sensory properties of the stimuli influenced inhibitory performance [17]. This influence was associated with activation differences in the cortical network known to underlie inhibitory control, highlighting the interactions between low-level sensory content and high-level control of action processes [17, 18]. Importantly, a redundant signal effect has been found to improve the inhibitory performances when presenting two congruent visual stop signals as compared to the use of a single visual stop signal [19]. Still, such redundant signal effect has not been investigated with bimodal or ecological stimulations, and it remains unknown whether audiovisual integration is beneficial not only for initiating action, but also for inhibiting one. Indeed, providing that a multimodal stop signal leads to a redundant signal effect, the stop process's resulting shortening would, in the context of the stop-signal paradigm, generate lower response probability and shorter SSRT than unimodal stop signals. Such a hypothesis is based on the widely accepted model of a convergence of the different modalities leading to an integration mechanism that occurs at the low level of information processing in early sensory areas. Numerous studies have described a shortening of sensory responses of single cells responses [20] and an influence on motor processing speed [21]. Indeed, a recent study has demonstrated that multisensory integration accelerates neural processing at the sensory encoding stages as well as during decision formation [22]. In

contrast, poorer inhibitory performances obtained with a multimodal stop signal, compared to unimodal ones would indicate either a failure of the a multisensory integration in the stop-signal processing or a strong regulation of multisensory integration through top-down control [23]. Such absence of multisensory benefits would suggest that response inhibition mechanisms could be segregated from those of action, adding further evidence that the cortical mechanisms of multisensory integration are clearly dependent on the behavioral goal.

In the present study, we compared the efficiency of auditory, visual, and audiovisual stop signals using ecological communication stimuli. We hypothesized that as multisensory integration leads to shorter RTs, the same should apply in a stop-signal paradigm, with an audiovisual stop signal leading to shorter SSRTs and lower response probabilities than a unimodal one. We tested our results against the independent race model to ensure that SSRTs were correctly simulated. In addition, we assessed the impact of ecological semantic load on action inhibition. The semantic content of stimuli, especially human faces, is known to enhance accuracy and RTs in detection and categorization tasks [24]. Similarly, the human voice is considered to be a specific communication signal [25], and vocal stimuli, like faces, are processed in specific cortical regions [26]. To test the semantic load effect, we used both natural and degraded stimuli in a classic stop-signal paradigm.

Materials and methods

Participants

Participants were 30 healthy volunteers (see below): 10 (six women) participated in the visual experiment; 10 (five women) in the auditory experiment; and 10 (five women) in the audiovisual experiment. Participants were aged 19–30 years ($M = 23$ years). They had normal or corrected-to-normal vision, and no hearing problems were reported. Participants provided sociodemographic characteristics and written informed consent. The study was conducted according to the principles stated in the Declaration of Helsinki (2013) and was approved by the local research ethics committee (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale Toulouse II Avis N°2–03–34/Avis N°2).

Apparatus and stimuli

Visual stimuli were displayed on a computer screen (19") with a refresh rate of 91 Hz. The computer was equipped with the Serial Response Box™ (SRB), which featured five buttons. One button was used to record the RT. The SRB was placed on the desk in front of the computer. The stimuli were presented using E-Prime software.

In the main task, the visual stimuli were static black-and-white images of different categories (animals, objects, nature scenes [27], except for faces). The images were normalized for intensity, luminance, and contrast. The same set of images was used as visual (only) stimuli for the three experiments. The only difference between groups was the stop-signal modality used after the main task visual stimulus.

In the visual modality, the stop signal was a static neutral female face in the ecological condition and a degraded face (using Fourier phase randomization) in the other condition. In the auditory modality, the stop signal was a female voice saying "Bah" in the ecological condition. In the other condition, the same vocal stimulus was degraded by the 2-band vocoder to create the stop signal [28]. In the audiovisual modality, the stop signal in the ecological condition was a combination of the static neutral female face and the female voice saying "Bah", and in the other condition, the face and the "Bah" sound were both degraded, as described above. A static image was used under the assumption that using a dynamic face would be crucial only for the

realization of tasks of recognition of emotions present on faces [29]. The stimuli onsets were closely aligned with a precision of 5 ms.

The stimuli used as stop signals were similar to the ones we used in previous studies based on the same detection task with the same protocols of RTs evaluation [30, 31]. When presented in a visuo–auditory modality, this type of signals induces a shortening of the RTs resulting from a multisensory integration as revealed by the violation of the race model. By accepting a Type 1 risk α of .05 and a statistical power $1 - \beta$ of .95 and an effect size .25 (number of groups 3, number of measures 8, correlations of repeated measures 0.5, non–sphericity correction 1) we can estimate the number of participants for the “ANOVA repeated measures within–between interactions” (using G*power 3.1.9.7 software [32]) as 30 participants, i.e., 10 participants per group. Moreover, in previous studies, SSRT values were convincingly compared between conditions when manipulating stimulation modality in 10 participants [33] or response modality in 9 participants [34].

Procedure

In each trial, a white fixation cross was followed by the main task image, for which participants had to press the response button as quickly as possible. In 20% of the trials (stop trials), the main task stimulus was followed by the stop signal requiring the participant to cancel the response. Presentation of the fixation cross varied between 750 and 1500 ms, to prevent automatic reactions. The main task stimulus remained on the screen for 80 ms. Stimulus onset asynchrony (SOA) between the end of the main task stimulus and the stop signal varied randomly (0 ms, 25 ms, 50 ms, 75 ms, or 100 ms). Participants were asked to focus on the main task image requiring quick responses. They were also instructed that in some stop trials, they would fail to cancel the response but that they should not be troubled by these failures.

Participants in each group (auditory, visual, and audiovisual) performed 200 trials in each condition (ecological vs. degraded stop signals). The session was divided into four experimental blocks (100 trials per block).

Results on response probabilities were set against the independent race model for stop signal tasks developed by Boucher et al. [13] (see S1 File), and SSRTs were calculated using the integration method devised by Verbruggen and Logan [35]. Briefly, for each SOA, the main stimulus RTs for no–stop trials were rank–ordered, and the n^{th} RT was selected, where n was the number of RTs multiplied by $p(\text{response}_{(\text{SOA})})$. The SOA was then subtracted to estimate the SSRT. Estimated SSRTs for different SOAs were then averaged to obtain a single SSRT for each participant, by condition.

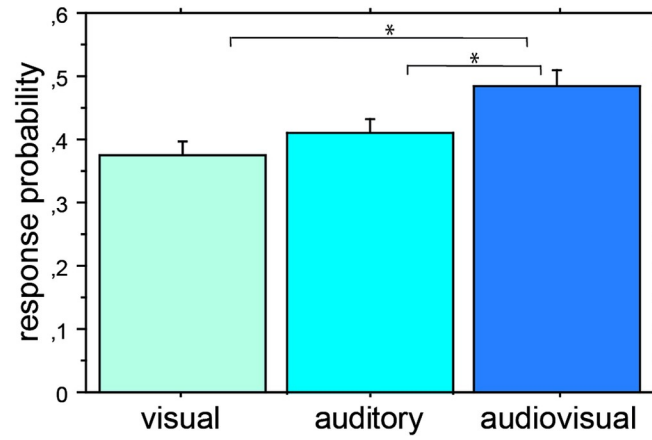
Results

Effects of stimulation modality

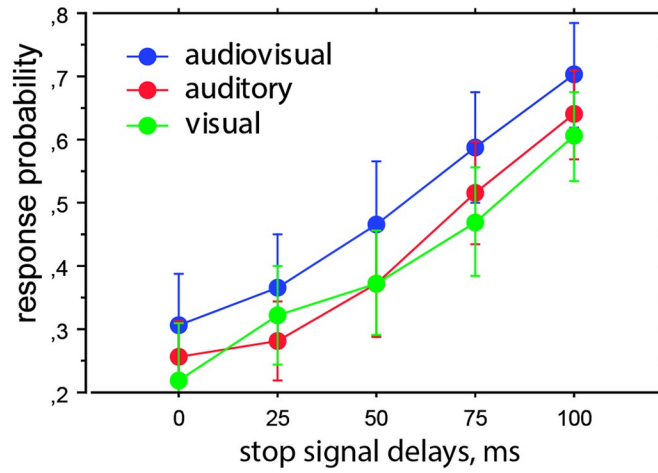
Response probabilities. We began by assessing the efficiency of each sensory modality in terms of response probability in the stop–signal task. Given that the distribution of response probabilities differed significantly from normal (Kolmogorov–Smirnov test; $p = .021$), we used nonparametric statistics to assess the effect of stimulation modality on the probability of responding.

The effect of stop signal modality, as reflected by response probability, could be assessed either overall or with respect to the various SOAs between the main task stimulus and the stop signal. For the former, the Kruskal–Wallis test showed a significant effect of modality on response probability ($H = 7.78$, $p = .021$, $df = 2$), which was significantly higher in the audiovisual modality than in either the auditory (Mann–Whitney test, $U = 4087.5$, $z = -2.23$, $p = .025$, $f = .41$) or visual (Mann–Whitney test, $U = 3959.5$, $z = -2.54$, $p = .011$, $f = 0.40$; Fig 1A)

A.



B.



C.

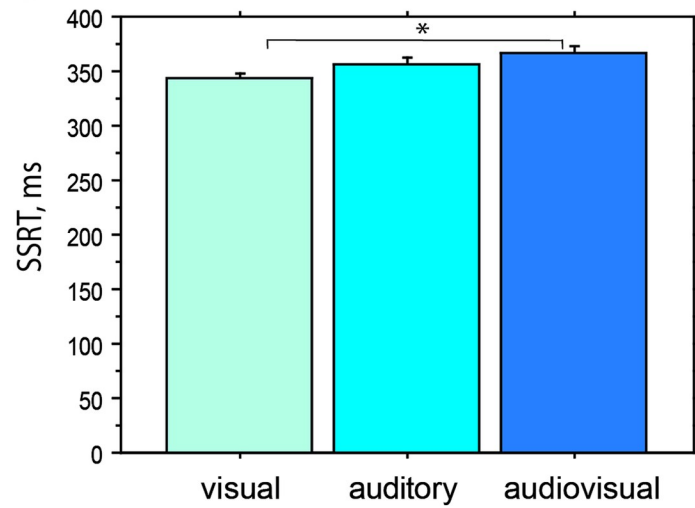


Fig 1. Effect of modality on response probability and stop-signal reaction time. *Note.* Panel A indicates the effect of stimulation modality on response probability after the stop signal. Response probability was significantly lower (meaning more efficient inhibition) for the visual and auditory stop signals than for the audiovisual stop signals. Panel B illustrates the changes in response probability as a function of SOA duration. Across all modalities, response probability increased with the stop signal delay. Panel C shows the effects of stimulation modality on SSRTs. SSRTs were significantly longer for audiovisual stop signals than for visual ones. * $p < .05$.

<https://doi.org/10.1371/journal.pone.0251739.g001>

modalities. The multisensory stop signal was, therefore, statistically less efficient than the unimodal ones. No significant difference in response probability was found between the visual and auditory modalities of the stop signal (Mann–Whitney test, $U = 4964.5$, $z = -1.23$, $p = .223$).

Concerning the analysis of the different SOAs, as classically reported in stop-signal experiments [14], we observed that the probability of canceling an action decreased when the interval between the presentation of the main task stimulus and the stop signal increased. Response probability significantly depended on the stop signal delay (Kruskal–Wallis test, $p = .0001$; Fig 1B), insofar as it was easier to cancel the response for short SOAs. This was observed in the visual (Kruskal–Wallis test $H = 35.08$, $p = .0017$), auditory (Kruskal–Wallis test, $H = 47.49$, $p = .0001$), and audiovisual (Kruskal–Wallis test, $H = 36.97$, $p = .0007$) modalities.

Stop-signal reaction times. We verified that the main task stimulus RTs for the failed stop trials were shorter on average than RTs for the no-stop trials (t -test, $p = .034$), thus validating the race model to compute SSRTs with our data [13, 14]. See also S1 File for the correspondence of our results with the race model of Boucher et al. [13]. As SSRT distribution did not differ significantly from normal (Kolmogorov–Smirnov test; $p = .22$), we performed an analysis of variance (ANOVA) with original stimulation modality as a factor and observed a significant effect of modality on SSRT, $F(2, 147) = 3.5$, $p = .032$. A post hoc Fisher comparison revealed that the mean SSRT was longer for the audiovisual modality ($M = 367$ ms) than for the visual modality ($M = 344$ ms), ($p = .009$, $d = .38$; Fig 1C). This difference reinforced the results on response probability and demonstrated that audiovisual stop signals were less efficient than visual or auditory ones. In order to further validate that this difference was imputable to inhibitory processes, we compared RTs to the main task image between the visual ($M = 428$ ms) and audiovisual ($M = 430$ ms) modalities. No difference was found (Fisher test, $p = .42$, $d = .02$), indicating that participants were similarly engaged in the main task, independently of the group modality.

Effects of stimulation degradation

Response probabilities. We further explored the impact of semantic content on the efficiency of the stop signal by comparing the original signals with their degraded counterparts. When we distinguished between the two conditions, we observed an effect of sensory modality on response probabilities with the original stop signals (Kruskal–Wallis test, $H = 6.31$, $p = .043$, $df = 2$). In particular, the original audiovisual stop signal had a higher response probability than the visual modality ($p = .0014$), but the difference between the original audiovisual and auditory modalities only tended toward significance ($U = 1013.5$, $z = -1.63$, $p = .10$). We failed to find an effect of sensory modality with the degraded stop signals (Kruskal–Wallis test; $H = 2.59$, $p = .275$, $df = 2$).

We then compared the original and degraded stop signals within each sensory modality. Using the paired Wilcoxon test for response probabilities, we showed that the original faces were a better stop signal than their degraded counterparts in the visual modality ($p = .0002$, $d = .43$, $z = -3.72$) (Fig 2A). This difference between the two did not depend on the SOA (Kruskal–Wallis test, $p = .08$), as the probability of responding was significantly higher for the

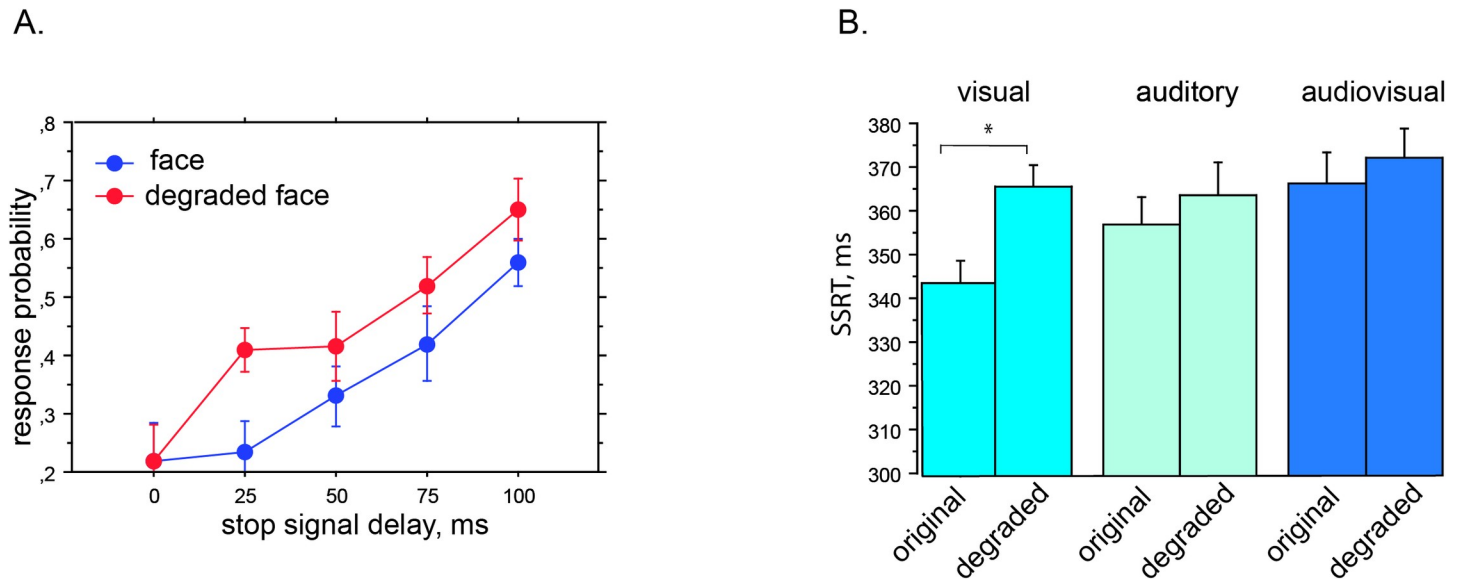


Fig 2. Effects of semantic load on response probability and stop-signal reaction time. Note. Panel A shows that response probability was lower for the original faces used as stop signals than for the degraded faces. This means that the original faces were more efficient in inhibiting action. Panel B indicates the effects of stop-signal modality and degradation on SSRTs. The effect of degradation on SSRTs was significant in the visual modality, as SSRTs were shorter for the original faces. * $p < .05$.

<https://doi.org/10.1371/journal.pone.0251739.g002>

degraded faces regardless of duration (Fig 2A). There was no significant difference in response probabilities between original voices and degraded voices in the auditory modality (paired Wilcoxon test; $p = .48$, $z = -0.71$). Similarly, there was no significant difference in response probabilities between the original and degraded stop signals in the audiovisual modality (paired Wilcoxon test; $p = .19$, $z = -1.31$), meaning that inhibition capacity did not depend on the semantic content of the audiovisual stop signal.

Stop-signal reaction times. The ANOVA with the condition factor, $F(5, 294) = 2.6$, $p = .026$ (Fig 2B), revealed that there was only a significant difference in SSRT between the original and degraded versions of the stop signals in the visual modality (post hoc Fisher test; $p = .013$, $d = .62$). In the two-way ANOVA with the modality (visual, auditory, audiovisual) and semantic content (original, degraded) factors, semantic content had a significant effect, $F(1, 294) = 5.2$, $p = .024$, but the effect of modality only tended toward significance ($p = .06$), $F(2, 294) = 2.8$. The interaction between modality and semantic content was not significant, $F(2, 294) = 1.1$, $p = .35$. As indicated above, the modality effect was significant in the one-way ANOVA. A post hoc Fisher test after the two-way ANOVA indicated a significant difference between the audiovisual and visual modalities ($p = .02$), confirming the above-described result in the one-way ANOVA.

Discussion

Our results indicate that, contrary to our initial hypothesis, a multisensory stop signal is much less efficient than a unimodal one in stopping an action. Response probabilities were higher for audiovisual stop signals than for unimodal ones. We found a similar contradiction of our initial hypothesis for the SSRTs calculated on the basis of the race model, as SSRTs were also longer for audiovisual stop signals. Further, semantic load only had a significant effect for visual stop signals, where the original faces were better stop signals than the degraded ones. No such effect of semantic load was observed for the audiovisual stop signals.

The most important result concerned the increase in response probabilities for audiovisual stop signals, compared with unimodal ones. We had hypothesized that multisensory facilitation (i.e., shorter RTs for audiovisual stimuli) would similarly lead to greater inhibition of motor responses, but results showed that audiovisual stop signals failed to provide a high inhibition level. On the contrary, the highest inhibition level was observed for unimodal stop signals, with lower response probabilities than those for the audiovisual modality. These results disproved our initial hypothesis but were in line with previous studies demonstrating poorer performances with multisensory signals in tasks involving executive control of action. For instance, in a go/no-go experiment, the rate of false alarms in no-go trials was higher in the visuotactile condition than in the visual one [36]. It is impossible to directly compare RTs for classic paradigms and stop-signal ones. However, SSRTs can be estimated using the independent race model for stop-signal tasks [13]. Our observation of lower inhibition levels for multisensory signals was corroborated by the SSRT analysis, as SSRTs were longer for bimodal stimulation, whereas in classic paradigms, RTs for bimodal stimulation are shorter.

To understand the *audiovisual impairment* demonstrated by our results, we can note that in most of the usual detection tasks showing a multisensory facilitation effect, the participants are cognitively free, ready to perform the task, and attentive to the upcoming stimulus. In contrast to this optimum situation, our participants were engaged in the main task when the stop signal occurred. This infrequent stimulation created a conflict in the monitoring of the main task, in which the participants had to stop what they were doing [37, 38]. In this particular situation of mental workload, we suggest that the additional sensory modality used for the bimodal stop signal increases the level of the conflict and, consequently, the task's cognitive demand, leading to decreased efficiency in stopping the action. Instead of being processed as a unified multisensory object, as is proposed for audiovisual speech [39], the face-voice stop stimulus is processed as multiple concurrent stop signals, which slows down decision-making. However, as we clearly showed that semantic content is important (see [Discussion](#) below), the multisensory interaction in an inhibition task may depend on the type and combination of sensory modalities [36].

At the cortical level, audiovisual integration is known to modulate brain activity in multiple cortical sites, including low- and high-level networks [40]. The multisensory integration that takes place in the early stages of sensory processing is subsequently processed in motor-related cortical or subcortical areas and can enhance response outcomes [20, 21]. However, during a decisional conflict, multisensory integration modulates brain activity at higher cortical levels by increasing the cognitive effort needed to modify the response. Brain activity related to the less efficient response in the audiovisual modality has been associated with an increase in the P300 event-related potential, compared with activity related to more efficient responses in unimodal modalities [41]. This has been interpreted as reflecting the greater cognitive effort of simultaneously processing multiple inputs from different modalities. Thus, we may have observed higher response probabilities and longer SSRTs in the audiovisual modality because of heightened task demand. Participants had to monitor the conflict generated by the initiation of the response to the main task and the occurrence of an unexpected stop signal (which occurred in a minority of the trials). In this context, the concurrent information added by the bimodality of the signal may have impaired the simultaneous execution of central operations (i.e., responding to the main task stimulus, monitoring the conflict with the stop signal, inhibiting the response). Modality pairing may compete with the task demand and constitute a limitation in the form of a central bottleneck in the context of multitasking [42]. Loss of executive control in a decisional situation has been linked to perseveration errors in human operators [43]. There is also now evidence that unexpected salient stimuli can fail to reach awareness when individuals are engaged in cognitively demanding situations [44]. To sum up, when

individuals are engaged in high-level conflict (e.g., response revision), multisensory signal presentation generates a high cognitive load. This increase in the supplementary task demand may impair action inhibition mechanisms.

In contrast to the inhibitory ineptness of the audiovisual stop signal in our study, the facial stop signal turned out to most efficiently inhibit action, reflected in the lower response probability and shorter SSRT. In the stop-signal paradigm, auditory signals are classically found to enhance both the speed and efficiency of response inhibition more than visual signals do [16, 45]. In our experiment, the facial stop signals in the visual modality led to better inhibition than the signals in the auditory and audiovisual modalities. This enhanced inhibition appears to have been face-specific, as it disappeared for the degraded faces with preserved spatial frequencies. *Face-specificity* means that inhibition is not merely mechanically triggered by a visual signal but that it is sensitive to the semantic content, which needs to be behaviorally meaningful to stop action. To the best of our knowledge, the face effect in visual stop-signal tasks had not previously been investigated, although the visual salience of the stop signal was already known to affect movement suppression, leading to shorter SSRTs and lower response probabilities [46, 47]. Faces constitute particularly salient stimuli, which we can detect and recognize significantly faster than nonfacial stimuli [48]. Moreover, a study using the anti-saccade paradigm showed that people have difficulty avoiding saccading to and fixating facial stimuli [49]. Thus, the salience of faces as stop signals renders inhibition faster and more accurate than with auditory stop signals. However, in our study, the visual face modality lost its efficacy to trigger the response inhibition process when it was combined with the auditory modality. The face facilitation seems not, therefore, to have been sufficient to offset the cost incurred by the increased task demand when concurrent multimodal information was added.

According to the standard horse-race model [12], the response initiation process (triggered by the main task stimulus) competes versus the response inhibition process (triggered by the stop signal). Thus, the result of the race determines whether the response is successfully canceled or not. The two processes are presumed to be stochastically independent (i.e., the finishing time of the response initiation process is independent of the finishing time of the response inhibition process in a given trial). However, the race model does not require functional independence between the two processes. A *functional* dependence would occur, providing a common factor was in capacity to influence both the duration of the response initiation and inhibition processes. Although this functional independence is still under debate [50], our results do not argue in favor of a functional linkage between the two processes. The multisensory content of a stimulation, known to facilitate the response to a go stimulus (see [Introduction](#)), appeared to damage the response to a stop signal. This impairment can probably be attributed to the inhibitory characteristic of the stop signal, rather than its unexpectedness. For instance, rare audiovisual stimuli were easier to detect than visual ones in an oddball task [51, 52]. This argues in favor of an audiovisual impairment in the perception of the stop signal, owing to its contradiction with the initiated main task response. One future avenue worth pursuing would be to test main task stimuli in different modalities (visual, auditory, audiovisual) to look for the possible interaction between the modality of the main task and that of the stop signal. It would also be worthwhile exploring the brain responses evoked by different modalities of the primary stimulus and stop signal within the same task to gain a better understanding of the multisensory facilitation and multisensory impairment processing stages.

Furthermore, individuals with attention deficit hyperactivity disorder exhibit both poorer inhibition of action [53, 54] and abnormal audiovisual integration, compared with neurotypical controls [55, 56]. Their inhibitory performances with visual stop signals could thus shed light on the interaction between cognitive control and modality pairing. The audiovisual impairment observed in the stop paradigm also suggests that redundant sensory information

increases cognitive load and negatively impacts decision-making, a result that may apply to irrational actions or strategies observed among the operators of critical systems (e.g., aircraft pilots, automobile drivers) when confronted with multiple warning signals [43].

Conclusions

Multisensory integration has been the subject of considerable research but has not yet been investigated in a situation calling for the executive control of thought and action. The present experiment makes it clear that modality pairing can have a dramatic effect on inhibitory control capabilities. Whereas visual (faces) stop signals were found to enhance inhibition, audiovisual stop signals led to higher response probabilities and longer SSRTs. We hypothesize that when people are engaged in an executive task, multimodal signals are not processed as a simple or unified object but as multiple sources of information. Concurrent information resulting from modality pairing in such a decisional situation may increase task demand and generate cognitive load, thereby impairing behavioral performance. However, an alternative explanation could be that subjects cannot disengage from the main active visual task's attentional process. The specific for multisensory integration top-down attentional control mechanism [23] could be too late to engage a complex fronto-parietal network that would allow stopping the action. This delay could lead to momentary multisensory inattentive phenomena that result in perseveration behavior [43]. To the best of our knowledge, this is the first demonstration of an audiovisual impairment that contradicts the theoretical models (e.g., the co-activation model), which describe multisensory integration as a facilitator in the domain of perception and action.

Supporting information

S1 File. Theoretical inhibition function and response probabilities.
(PDF)

Acknowledgments

The authors wish to thank all participants in this research.

Author Contributions

Conceptualization: Kuzma Strelnikov, Pascal Barone.

Data curation: Kuzma Strelnikov, Mario Hervault, Lidwine Laurent.

Formal analysis: Kuzma Strelnikov, Mario Hervault, Lidwine Laurent.

Funding acquisition: Mario Hervault, Pascal Barone.

Investigation: Kuzma Strelnikov, Lidwine Laurent.

Methodology: Kuzma Strelnikov, Mario Hervault, Lidwine Laurent, Pascal Barone.

Project administration: Pascal Barone.

Resources: Pascal Barone.

Supervision: Pascal Barone.

Validation: Mario Hervault, Pascal Barone.

Writing – original draft: Kuzma Strelnikov, Mario Hervault, Pascal Barone.

Writing – review & editing: Kuzma Strelnikov, Mario Hervault, Pascal Barone.

References

1. Sumbly WH, Pollack I. Visual contribution to speech intelligibility in noise. *J Acoust Soc Am*. 1954; 26: 212–215. <https://doi.org/10.1121/1.1907309>
2. Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex N Y N* 1991. 2007; 17: 1147–1153. <https://doi.org/10.1093/cercor/bhl024> PMID: 16785256
3. Cappe C, Rouiller EM, Barone P. Multisensory anatomical pathways. *Hear Res*. 2009; 258: 28–36. <https://doi.org/10.1016/j.heares.2009.04.017> PMID: 19410641
4. Miller J. Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Atten Percept Psychophys*. 2016; 78: 516–519. <https://doi.org/10.3758/s13414-015-1017-z> PMID: 26555650
5. Raab DH. Statistical facilitation of simple reaction times. *Trans N Y Acad Sci*. 1962; 24: 574–590. <https://doi.org/10.1111/j.2164-0947.1962.tb01433.x> PMID: 14489538
6. Hershenson M. Reaction time as a measure of intersensory facilitation. *J Exp Psychol*. 1962; 63: 289–293. <https://doi.org/10.1037/h0039516> PMID: 13906889
7. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory–visual interactions during early sensory processing in humans: a high–density electrical mapping study. *Cogn Brain Res*. 2002; 14: 115–128. [https://doi.org/10.1016/s0926-6410\(02\)00066-6](https://doi.org/10.1016/s0926-6410(02)00066-6) PMID: 12063135
8. Miller J. Divided attention: Evidence for coactivation with redundant signals. *Cognit Psychol*. 1982; 14: 247–279. [https://doi.org/10.1016/0010-0285\(82\)90010-x](https://doi.org/10.1016/0010-0285(82)90010-x) PMID: 7083803
9. Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J Exp Psychol Hum Percept Perform*. 2012; 38: 1517–1529. <https://doi.org/10.1037/a0027339> PMID: 22390292
10. Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT. Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res*. 2004; 158: 405–414. <https://doi.org/10.1007/s00221-004-1913-2> PMID: 15221173
11. Stein BE. *The New Handbook of Multisensory Processing*. MIT Press; 2012.
12. Logan GD, Cowan WB. On the ability to inhibit thought and action: A theory of an act of control. *Psychol Rev*. 1984; 91: 295–327. <https://doi.org/10.1037/0033-295X.91.3.295>
13. Boucher L, Palmeri TJ, Logan GD, Schall JD. Inhibitory control in mind and brain: an interactive race model of countermanding saccades. *Psychol Rev*. 2007; 114: 376–397. <https://doi.org/10.1037/0033-295X.114.2.376> PMID: 17500631
14. Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop–signal task. Frank MJ, Badre D, Egner T, Swick D, editors. *eLife*. 2019; 8: e46323. <https://doi.org/10.7554/eLife.46323> PMID: 31033438
15. Bari A, Robbins TW. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog Neurobiol*. 2013; 108: 44–79. <https://doi.org/10.1016/j.pneurobio.2013.06.005> PMID: 23856628
16. van der Schoot M, Licht R, Horsley TM, Sergeant JA. Effects of stop signal modality, stop signal intensity and tracking method on inhibitory performance as determined by use of the stop signal paradigm. *Scand J Psychol*. 2005; 46: 331–341. <https://doi.org/10.1111/j.1467-9450.2005.00463.x> PMID: 16014077
17. Friedrich J, Mückschel M, Beste C. Physical intensity of stimuli modulates motor inhibition by affecting response selection processes in right inferior frontal regions. *Behav Brain Res*. 2019; 359: 597–608. <https://doi.org/10.1016/j.bbr.2018.10.006> PMID: 30292901
18. Friedrich J, Beste C. The impact of stimulus modality on the processing of conflicting sensory information during response inhibition. *Neuroscience*. 2019; 410: 191–201. <https://doi.org/10.1016/j.neuroscience.2019.05.010> PMID: 31100340
19. Cavina-Pratesi C, Bricolo E, Prior M, Marzi CA. Redundancy gain in the stop–signal paradigm: implications for the locus of coactivation in simple reaction time. *J Exp Psychol Hum Percept Perform*. 2001; 27: 932–941. PMID: 11518154
20. Wang Y, Celebrini S, Trotter Y, Barone P. Visuo–auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci*. 2008; 9: 79. <https://doi.org/10.1186/1471-2202-9-79> PMID: 18699988
21. Miller J, Ulrich R, Lamarre Y. Locus of the redundant–signals effect in bimodal divided attention: a neurophysiological analysis. *Percept Psychophys*. 2001; 63: 555–562. <https://doi.org/10.3758/bf03194420> PMID: 11414141
22. Mercier MR, Cappe C. The interplay between multisensory integration and perceptual decision making. *NeuroImage*. 2020; 222: 116970. <https://doi.org/10.1016/j.neuroimage.2020.116970> PMID: 32454204

23. De Meo R, Murray MM, Clarke S, Matusz PJ. Top-down control and early multisensory processes: chicken vs. egg. *Front Integr Neurosci*. 2015; 9. <https://doi.org/10.3389/fnint.2015.00017> PMID: 25784863
24. Tsao DY, Livingstone MS. Mechanisms of face perception. *Annu Rev Neurosci*. 2008; 31: 411–437. <https://doi.org/10.1146/annurev.neuro.30.051606.094238> PMID: 18558862
25. Belin P, Bestelmeyer PEG, Latinus M, Watson R. Understanding voice perception. *Br J Psychol Lond Engl* 1953. 2011; 102: 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x> PMID: 21988380
26. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000; 403: 309–312. <https://doi.org/10.1038/35002078> PMID: 10659849
27. Juan C, Cappe C, Alric B, Roby B, Gilardeau S, Barone P, et al. The variability of multisensory processes of natural stimuli in human and non-human primates in a detection task. *PLOS ONE*. 2017; 12: e0172480. <https://doi.org/10.1371/journal.pone.0172480> PMID: 28212416
28. Rouger J, Lagleyre S, Fraysse B, Deneve S, Deguine O, Barone P. Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci U S A*. 2007; 104: 7295–7300. <https://doi.org/10.1073/pnas.0609419104> PMID: 17404220
29. Watson R, Latinus M, Noguchi T, Garrod OGB, Crabbe F, Belin P. Dissociating task difficulty from incongruence in face-voice emotion integration. *Front Hum Neurosci*. 2013; 7. <https://doi.org/10.3389/fnhum.2013.00744> PMID: 24294196
30. Salles J, Strelnikov K, Carine M, Denise T, Laurier V, Molinas C, et al. Deficits in voice and multisensory processing in patients with Prader-Willi syndrome. *Neuropsychologia*. 2016; 85: 137–147. <https://doi.org/10.1016/j.neuropsychologia.2016.03.015> PMID: 26994593
31. Debladis J, Valette M, Strelnikov K, Mantoulan C, Thuilleaux D, Laurier V, et al. Face processing and exploration of social signals in Prader-Willi syndrome: a genetic signature. *Orphanet J Rare Dis*. 2019; 14: 262. <https://doi.org/10.1186/s13023-019-1221-3> PMID: 31730500
32. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009; 41: 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149> PMID: 19897823
33. Mirabella G, Pani P, Paré M, Ferraina S. Inhibitory control of reaching movements in humans. *Exp Brain Res*. 2006; 174: 240–255. <https://doi.org/10.1007/s00221-006-0456-0> PMID: 16636792
34. Brunamonti E, Ferraina S, Paré M. Controlled movement processing: evidence for a common inhibitory control of finger, wrist, and arm movements. *Neuroscience*. 2012; 215: 69–78. <https://doi.org/10.1016/j.neuroscience.2012.04.051> PMID: 22554783
35. Verbruggen F, Logan GD. Models of Response Inhibition in the Stop-Signal and Stop-Change Paradigms. *Neurosci Biobehav Rev*. 2009; 33: 647–661. <https://doi.org/10.1016/j.neubiorev.2008.08.014> PMID: 18822313
36. Bodmer B, Beste C. On the dependence of response inhibition processes on sensory modality. *Hum Brain Mapp*. 2017; 38: 1941–1951. <https://doi.org/10.1002/hbm.23495> PMID: 28045223
37. Braver TS, Barch DM, Gray JR, Molfese DL, Snyder A. Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb Cortex N Y N* 1991. 2001; 11: 825–836. <https://doi.org/10.1093/cercor/11.9.825> PMID: 11532888
38. Mirabella G. Should I stay or should I go? Conceptual underpinnings of goal-directed actions. *Front Syst Neurosci*. 2014; 8: 6. <https://doi.org/10.3389/fnsys.2014.00006> PMID: 24574979
39. Vatakis A, Ghazanfar AA, Spence C. Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *J Vis*. 2008; 8: 14.1–11. <https://doi.org/10.1167/8.9.14> PMID: 18831650
40. Murray MM, Spierer L. Multisensory Integration: What You See Is Where You Hear. *Curr Biol*. 2011; 21: R229–R231. <https://doi.org/10.1016/j.cub.2011.01.064> PMID: 21419991
41. Stock A-K, Gohil K, Huster RJ, Beste C. On the effects of multimodal information integration in multi-tasking. *Sci Rep*. 2017; 7: 1–12. <https://doi.org/10.1038/s41598-016-0028-x> PMID: 28127051
42. Hazeltine E, Ruthruff E. Modality pairing effects and the response selection bottleneck. *Psychol Res*. 2006; 70: 504–513. <https://doi.org/10.1007/s00426-005-0017-3> PMID: 16151721
43. Dehais F, Hodgetts HM, Causse M, Behrend J, Durantin G, Tremblay S. Momentary lapse of control: A cognitive continuum approach to understanding and mitigating perseveration in human error. *Neurosci Biobehav Rev*. 2019; 100: 252–262. <https://doi.org/10.1016/j.neubiorev.2019.03.006> PMID: 30878500
44. Causse M, Imbert J-P, Giraudet L, Jouffrais C, Tremblay S. The Role of Cognitive and Perceptual Loads in Inattentional Deafness. *Front Hum Neurosci*. 2016; 10. <https://doi.org/10.3389/fnhum.2016.00344> PMID: 27458362

45. Ramautar JR, Kok A, Ridderinkhof KR. Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. *Biol Psychol.* 2006; 72: 96–109. <https://doi.org/10.1016/j.biopsycho.2005.08.001> PMID: 16157441
46. Montanari R, Giamundo M, Brunamonti E, Ferraina S, Pani P. Visual salience of the stop-signal affects movement suppression process. *Exp Brain Res.* 2017; 235: 2203–2214. <https://doi.org/10.1007/s00221-017-4961-0> PMID: 28439628
47. Pani P, Giarrocco F, Giamundo M, Montanari R, Brunamonti E, Ferraina S. Visual salience of the stop signal affects the neuronal dynamics of controlled inhibition. *Sci Rep.* 2018; 8: 1–13. <https://doi.org/10.1038/s41598-017-17765-5> PMID: 29311619
48. Palermo R, Rhodes G. Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia.* 2007; 45: 75–92. <https://doi.org/10.1016/j.neuropsychologia.2006.04.025> PMID: 16797607
49. Morand SM, Grosbras M–H, Caldara R, Harvey M. Looking away from faces: influence of high-level visual processes on saccade programming. *J Vis.* 2010; 10: 16.1–10. <https://doi.org/10.1167/10.3.16> PMID: 20377293
50. Matzke D, Verbruggen F, Logan GD. The Stop-Signal Paradigm. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience.* American Cancer Society; 2018. pp. 1–45. <https://doi.org/10.1002/9781119170174.epcn510>
51. Campanella S, Bruyer R, Froidbise S, Rossignol M, Joassin F, Kornreich C, et al. Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol.* 2010; 121: 1855–1862. <https://doi.org/10.1016/j.clinph.2010.04.004> PMID: 20434394
52. Campanella S, Delle-Vigne D, Kornreich C, Verbanck P. Greater sensitivity of the P300 component to bimodal stimulation in an event-related potentials oddball task. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol.* 2012; 123: 937–946. <https://doi.org/10.1016/j.clinph.2011.10.041> PMID: 22119176
53. Bekker EM, Overtom CC, Kenemans JL, Kooij JJ, De Noord I, Buitelaar JK, et al. Stopping and changing in adults with ADHD. *Psychol Med.* 2005; 35: 807–816. <https://doi.org/10.1017/s0033291704003459> PMID: 15997601
54. Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? *J Abnorm Psychol.* 2005; 114: 216–222. <https://doi.org/10.1037/0021-843X.114.2.216> PMID: 15869352
55. Panagiotidi M, Overton PG, Stafford T. Multisensory integration and ADHD-like traits: Evidence for an abnormal temporal integration window in ADHD. *Acta Psychol (Amst).* 2017; 181: 10–17. <https://doi.org/10.1016/j.actpsy.2017.10.001> PMID: 29024843
56. McCracken HS, Murphy BA, Glazebrook CM, Burkitt JJ, Karellas AM, Yilder PC. Audiovisual Multisensory Integration and Evoked Potentials in Young Adults With and Without Attention-Deficit/Hyperactivity Disorder. *Front Hum Neurosci.* 2019; 13. <https://doi.org/10.3389/fnhum.2019.00095> PMID: 30941026