Using Signal Detection Theory to Investigate the Role of Visual Information in Performance Monitoring in Typing

Svetlana Pinet (s.pinet@bcbl.eu)
Basque Center on Cognition, Brain and Language
Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia-San Sebastian, Spain

Nazbanou Nozari (bnozari@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University,
5000 Forbes Ave. Pittsburgh, PA 15213 USA

Abstract

This paper uses the signal detection theory (SDT) to investigate the contribution of visual information to two monitoring-dependent functions, *metacognitive awareness of errors* and *error corrections*. Data from two experiments show that complete removal of visual outcome results in a mild decrease in error awareness and a much more significant decrease in correction rates. Partially restoring visual information by including positional information (as in masked password typing) causes a modest but statistically significant improvement in correction performance. Interestingly, participants treat the change to the quality of information differently across the tasks, with more conservative behavior (avoiding false alarms) in the correction task. These findings show the SDT's ability to quantify, in a graded manner, the contribution of specific types of information to monitoring in complex tasks, while also providing additional information about how participants handle the change to the quality of information in a task-dependent manner.

Keywords: language production; signal detection theory (SDT); monitoring; error awareness; typing.

Introduction

Monitoring refers to processes involved in the surveillance of one's cognitive and motor operations in order to ensure a satisfactory outcome. Complex cognitive operations, such as language production, are often carried out fluently and with few errors because of efficient underlying monitoring and control processes (Nozari, 2018). But what is the nature of such processes? Broadly speaking, there are two types of processes for action monitoring: *external* monitoring mechanisms are those that rely on the outcome of performance (with or without additional reliance on internally generated representations). *Internal* monitoring mechanisms, on the other hand, refer to mechanisms that rely entirely on internal representations to estimate the likelihood of errors before they become overt. Monitoring in language production is a good example for showcasing this dichotomy: people can hear what they say, or see what they write or type and judge its accuracy via auditory and visual systems, respectively (external-channel monitoring; e.g., Guenther, 2016). At the same time, a large body of past work has shown efficient monitoring when people do not have access to the outcome of the process, supporting alternative mechanisms that monitor performance internally (internal-channel monitoring; e.g., Hickok, 2012; Nozari et al., 2011). The question remains: How is labor divided between these two monitoring channels? This study uses signal detection theory (SDT) to answer this question.

Language production monitoring as SDT

The challenge of answering questions about the contribution of external vs. internal monitoring channels to monitoring stems from the fact that models of external and internal monitoring often propose very different mechanisms. To quantify performance across such models, a common currency is needed. A suitable candidate for such currency is the notion of *conflict* (Botvinick et al., 2001). In some monitoring models, conflict has been defined in a specific way. For example, Nozari et al. (2011) proposed an internal monitoring mechanism in which the probability of an error was estimated by detecting conflict between two representations (e.g., lexical representations of "cat" and "dog"). In this model, conflict is defined as the inverse of the difference between the activation of two representations within the same layer of the language processing system.

Other models do not resort to an explicit definition of conflict. Nevertheless, the concept of conflict can still be defined for such models. For example, Hickock (2012) proposed a mechanism in which a lexical item activates both its corresponding motor and sensory nodes. The former then suppresses the latter via the inhibitory connections. On error trials, the concept activates different motor and sensory nodes. Consequently, the sensory node, which could not be suppressed, remains active and generates an error signal. This account can be easily reframed in terms of conflict between two different layers of the language processing system (sensory and motor layers): during a correct response, the suppression of perceptual representation leads to low conflict between motor and perceptual representations. During an error trial, the high activation of the perceptual representation creates conflict with the motor representation.

Finally, accounts such as forward (and inverse) models that compare the sensory consequences of a motor command to the actual sensory input generated by the outcome (hence their categorization as external-channel models by our definition), effectively calculate conflict between predicted and sensory inputs within the same layer of the language processing system (Guenther, 2016). In short, the notion of conflict can be used more broadly to capture something essential about monitoring: regardless of the specific mechanism, the predictions of all the monitoring models can be quantified as *higher conflict (within or between layers of representation) on error compared to correct trials*.

The advantage of defining a common currency across different monitoring accounts is that a unified framework can be created to answer questions that are not specific to individual mechanisms, but instead encompass multiple monitoring mechanisms. Specifically, we are interested in whether —and to what extent— removing the outcome of production affects monitoring performance. This question, by definition, requires starting with a framework in which both external and internal monitoring mechanisms contribute to the process, and investigating the change to performance when the impact of one is reduced. Figure 1 shows this general framework in SDT terms. Production can have two possible outcomes: correct and error trials. Over many trials, two distributions emerge based on these response types. Importantly, the two distributions are, on average, associated with different values of conflict (see above for a model-independent definition of conflict). Monitoring can thus be viewed as a decision (albeit implicit) based on the amount of conflict. This decision, according to SDT, is determined by two factors: the distance between the two distributions, indexed by d', and the location of the criterion c (defined as the number of standard deviations from the position of the ideal observer, i.e., the location at the junction of the two distributions), above which responses will be labeled as errors and below which as correct (SDT-type 2, see Wixted, 2020, for a review). The value of d' is a function of task difficulty and state of the production system, while the criterion depends on participants' goals (Nozari & Hepner, 2019).

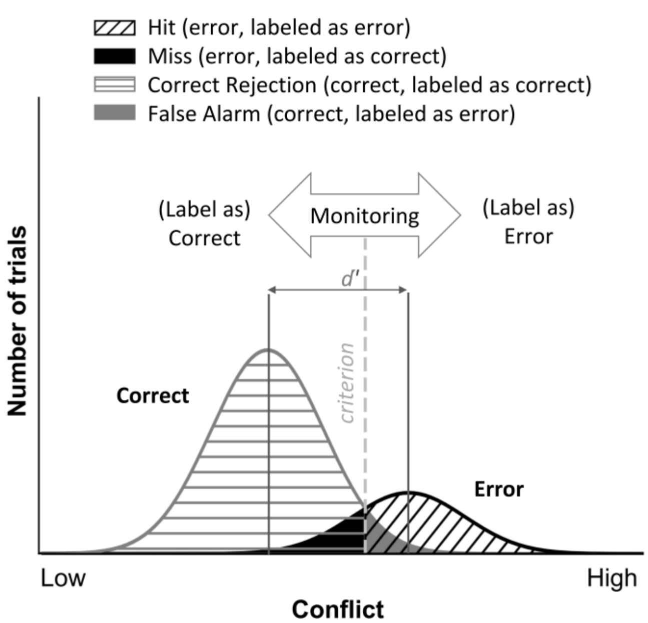

Figure 1. Monitoring in the SDT framework. Note that d' is directly dependent on the distance between the two distributions. Criterion, on the other hand, can be placed at any location.

In this framework, it is possible to investigate the effect of subtracting one kind of information (e.g., the outcome of performance or external input) on both d' and criterion. This, in turn, allows us to contrast different outcomes of monitoring. We target two such outcomes in the current study: *error awareness* (i.e., consciously reporting that an error has been committed), and *error correction* (attempting a repair). Since there is evidence that, at least in some tasks, repairs may be attempted in the absence of conscious awareness of errors, these two processes are potentially dissociable (e.g., Charles, Van Opstal, Marti, & Dehaene, 2013).

To investigate the effect of removing external information on error awareness and correction, we use a typing task, as an example of a language production task, for three reasons: a) past work has shown a strong influence of linguistic factors on typing, as well as a clear resemblance of typing errors to spoken errors, showing that typing is indeed a good example of a complex language production task (Pinet, Ziegler, & Alario, 2016; Nozari & Pinet, 2018). b) Unlike spoken production, in which the complete removal of the auditory signal via noise-masking is problematic because of bone conduction, the visual external input can be easily removed from typing. Finally, c) corrections in typing are clearly marked by the use of backspace, removing any subjectivity about whether a correction attempt has been made or not. In addition to these advantages, recent data show that monitoring shows the same general indices of monitoring found in action monitoring (Pinet & Nozari, 2020), making the findings of this study informative both for language production and for general action monitoring accounts.

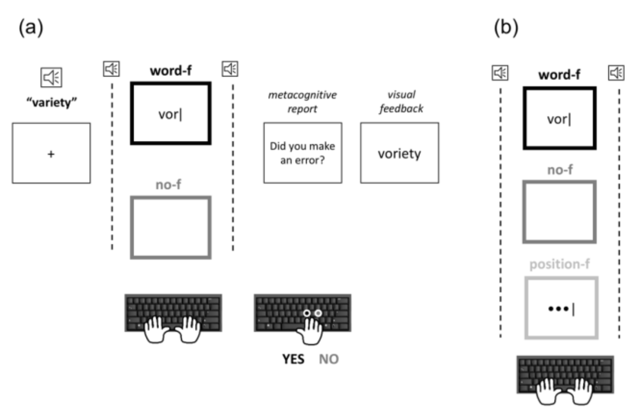

Figure 2. Trial structure in Exps 1 (a) and 2 (b). Exp 2 follows the same general structure as in Exp 1, with an additional condition. f = feedback.

In two experiments, we had people type the words they heard under a temporal deadline. In the baseline (word-feedback) condition, they saw what they typed on the screen in real time. In the no-feedback condition, the word did not appear on the screen until the end of the trial, with the goal of understanding the potential detriment to monitoring performance as a function of removing visual information. Based on prior results, we expected lower correction rates in the no-feedback compared to the word-feedback condition (Pinet & Nozari, 2020). The new manipulation in Exp 1 was adding a question right after typing, asking participants if they had made an error or not (metacognitive judgment; Fig. 2), as an independent measure of error awareness. Exp 2 aimed to replicate the findings of Exp 1 and further probe the specific role of visual information in monitoring. We added a new condition (position-feedback), in which participants saw dots (instead of letters) as they typed, similar to masked password typing (Fig. 2; position-f). This condition still provided visual information about the *position* of the letters, but not their *identity*, and could thus distinguish between the role of these two sources of visual information for monitoring.

The data from both experiments were coded as Hits, Misses, Correct Rejections, and False Alarms (see Fig. 1 for definitions) and were analyzed by SDT models which estimated *d'* and criterion parameters for each subject. Group analyses compared these parameters across conditions. A decrease in *d'* indexes the loss of information that cleanly teases apart the distributions of correct and error trials, and thus tells us about the necessity of certain kind of information for monitoring. A change in the criterion, on the other hand, tells us about how participants deal with the reduced quality of information. A shift to the left would indicate participants' desire to detect/correct as many errors as possible, even at the cost of erroneously marking some correct responses as errors, i.e., False Alarms. A shift to the right, on other hand, marks a tendency to avoid False Alarms, even if it means missing some errors. The comparisons, thus, paint a full picture of the contribution of visual information (positional vs. identity) to error awareness and corrections, as well as participants' strategies on how to use such information.

Experiment 1

Methods

Participants Participants were recruited via Amazon Mechanical Turk. Their eligibility to participate was determined using a short online task, in which they heard and typed words in two phases: in phase 1, they heard 15 words, one at a time, to type without time pressure. In phase 2, they heard another 15 words and had 2000 ms to finish typing them. A minimum criterion of 80% accuracy in phase 1, and 80% completed responses by the deadline with at least 50% accuracy in phase 2 was required for inclusion in the study. Forty-two native English-speaking participants (17 male; mean age= 35.6, SD = 8 years) who passed these criteria took part in the study in exchange for payment.

Stimuli Stimuli were 600 7- and 8-letter words from the English Lexicon Project database (Balota et al., 2007). Log-transformed word frequency ranged from 1.7 to 3. Plural forms, compound words, and words that had homophones, were not included. Words were divided into two lists of 300 items, balanced with regard to number of phonemes, syllables and letters, word and bigram frequency, and percentage of bimanual alternations. Words in both lists were recorded by a native English speaker.

Procedures The experiment was programmed using the jsPsych library (de Leeuw, 2015), embedded in an HTML environment, and completed online. The Python library psiTurk (Gureckis et al., 2016) was used to handle participants' recruitment and compensation.

The task was a typing-to-dictation task. On each trial, participants heard a word, followed by a beep (1000 Hz, 100 ms). They were instructed to finish typing the word before a second beep, 1800 ms after the first one (but their responses were registered for an additional 500 ms after the second beep). They could use backspace and correct their responses if they wished to. Immediately after that, participants completed the metacognitive judgment task, in which they answered yes/no to the question "Did you make an error in what you first typed?". Clear instructions were given that even corrected errors should be counted as errors (a "yes" response). Trials were separated by 500 ms intertrial intervals. There were breaks every 50 trials.

Participants completed two conditions in counterbalanced order, with list assignment to conditions also counterbalanced across participants. In the word-feedback (baseline) condition, participants saw, in real time, what they typed on the monitor. In the no-feedback condition, the screen remained blank as they typed. They were only shown the outcome of typing after answering the metacognitive judgment task (Fig 2a).

Analyses Performance measures included errors, response time (RTs), interkeystroke intervals (IKIs), and two measures of monitoring: metacognitive judgments (as measuring error awareness) and backspaces (as measuring corrections). Data were analyzed in R, using the psycho package version 0.4.91 (Makowski, 2018) for fitting signal detection models, and the lmerTests package for computing statistics (Kuznetsova, Brockhoff, & Christensen, 2017). Multilevel models with random effects of both subjects and items were used for the analyses of accuracy, RT and IKI data. Model-derived parameters must be calculated over all trials per subject, thus they were analyzed using linear regression models with random effect of subjects only. Error trials were excluded from RT and IKI analyses. RTs and IKIs shorter or longer that 3SD from participant means were also excluded.

Results & Discussion

One participant was excluded for failure to follow task instructions. The error rate in the no-feedback condition was lower than the word-feedback condition (21.2±10.3% vs. 22.6±10.7%; $\beta = 0.088$, $z = 2.6$, $p = .009$). Average RTs, on the other hand, were higher in the no-feedback compared to the word-feedback condition (390.2±76ms vs. 362±92ms; $\beta = 29.7$, $t = 16.8$, $p <.001$). The same was true for IKIs (168.7±24ms vs. 163.3±26ms; $\beta = 5.86$, $t = 15.2$, $p <.001$). Figure 3 shows the SDT measures for monitoring indices.

The overall rate of error awareness (hit rate) in metacognitive judgments was 69% and 54% in the word-feedback and no-feedback conditions, respectively. Model-estimated *d'* was significantly lower for the no-feedback compared to word-feedback condition (2.1 ± 0.4 vs. 2.9 ± 0.6; $\beta = -.81$, $t = -7.0$, $p <.001$), whereas the location of the criterion did not significantly differ between the two conditions (1.0 ± .4 vs. .9 ± .3; $\beta = .08$, $t = 1.1$, $p = .32$).

The overall rate of correction attempts (hit rate) was 28% (763 attempts) in the word-feedback and 8% (221 attempts) in the no-feedback conditions. Model-estimated *d'* was significantly lower for the no-feedback compared to word-feedback condition (1.2 ± .6 vs. 2.0 ± .8; $\beta = -.83$, $t = -5.1$, $p <.001$). Criterion was significantly higher in the no-feedback compared to the word-feedback condition (2.2 ± .4 vs. 1.7 ± .5; $\beta = .50$, $t = 5.1$, $p < .001$).

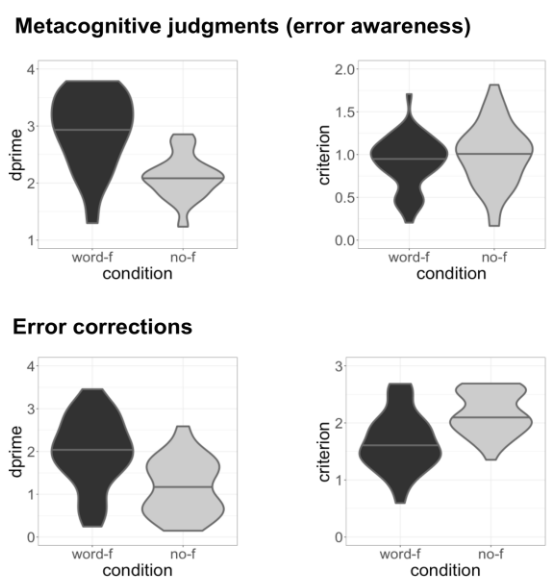

Figure 3. Results of metacognitive judgments (upper panel) and error corrections (lower panel) in Exp 1. Left panels show d-prime, and right panels criterion values.

To summarize, removing visual feedback slowed down typing, but participants were not any less accurate. There was a modest drop in error awareness and a much greater drop in correction attempts (22% vs. 71% change over baseline, respectively). Model-derived *d'* was significantly lower in the absence of visual information, for both error awareness and correction measures, showing that visual information contributed to cleaning up the conflict between error and correct responses in both tasks. Participants, however, treated the decreased quality of information differently in the two cases: while they did not shift the criterion position for metacognitive judgments, they significantly shifted it to the right for corrections, i.e., they sacrificed Hits in order to minimize False Alarms. This is most likely due to the fact that correction attempts are costly; participants must stop ongoing behavior, erase the mistake, replace the segment with a repair, and resume typing. This takes both time and effort. By shifting their criterion to the right, they minimize the chance of unnecessary corrections (False Alarms dropped from 3.5% to 1.6%). Yes/no metacognitive judgments, on the other hand, are not associated with similarly high cognitive costs, hence no rightward shift of the criterion despite lower *d'*s in this task.

The results of Exp 1 showed the importance of visual information for both conscious detection and correction of errors, with the latter taking a much bigger hit when visual information was absent. But which aspect of visual information is critical for repairs, the position of the error or the identity of the letters? Exp 2 answers this question.

Experiment 2

Methods

Participants Forty-two English speaking participants who had not participated in Exp 1 and who passed the screening test described under that experiment were recruited via Amazon Mechanical Turk and took part in the study in exchange for payment.

Stimuli The same materials as Exp 1 were used. The 600 words were divided into three lists, balanced for factors described under Exp 1.

Procedures Procedures were similar to Exp 1, except that each participant completed three conditions in counterbalanced order, with lists also counterbalanced with regard to condition across participants. Word-feedback and no-feedback conditions were identical to Exp 1. In the position-feedback condition instead of each letter a dot (•) appeared on the screen as participants typed their response, similar to password typing.

Analysis Analyses were similar to Exp 1. Contrasts were treatment-coded. To compare all our conditions to each other, we used two different contrasts coding schemes, alternatively taking word-feedback and no-feedback as the reference level.

Results & Discussion

One participant was excluded because of a technical issue in recording their keystrokes. The error rate was not significantly different between the word-feedback and no-feedback conditions (25.6 ± 13% vs. 24.9 ± 13%; β = 0.035, z = 0.86, p = .39), between the word-feedback and position-feedback (25.5 ± 12%; β = -0.005, z = -0.12, p = .91), or between the no-feedback and position-feedback (β = -0.042, z = -0.98, p =.33). RTs, on the other hand, were significantly slower for the no-word feedback compared to the word-feedback (377.3 ± 87ms vs. 359.2 ± 87ms; β = 19.6, t = 8.8, p <.001). They were also significantly slower for the position-feedback compared to the word-feedback condition (378.0 ± 84ms; β = 21.2, t = 9.6, p <.001). The comparison between the position-feedback and no-feedback revealed no significant difference (β = 1.6, t = 0.72, p = .48). IKIs were also significantly longer for the no-feedback compared to the word-feedback (178.0 ± 29ms vs. 172.1 ± 28ms; β = 6.1, t = 11.8, p < .001). The same was true for the comparison between position-feedback and word-feedback (176.0 ± 28ms; β = 4.2, t = 8.3, p <.001). The comparison between position-feedback and no-feedback revealed significantly longer IKI in the no-feedback condition (β = -1.8, t = -3.5, p <.001). Figure 4 shows the SDT measures for monitoring indices.

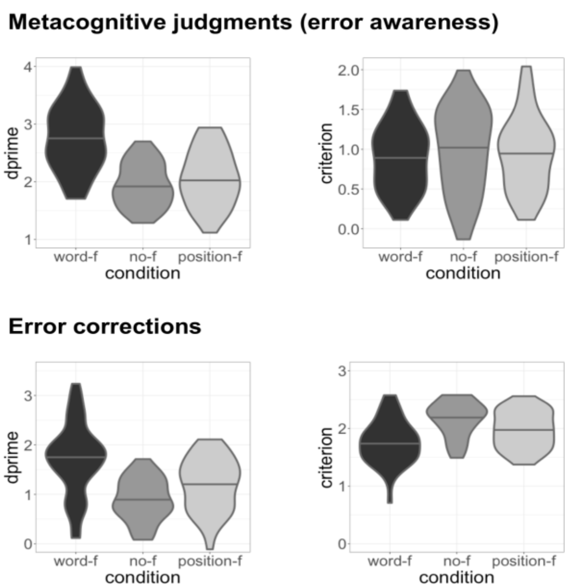

Figure 4. Results of metacognitive judgments (upper panel) and error corrections (lower panel) in Exp 2. Left panels show d-prime, and right panels criterion values.

The overall rate of error awareness (hit rate) in metacognitive judgments was 69% in the word-feedback, 55% in the no-feedback, and 57% in the position-feedback conditions, respectively. As in Exp 1, the average *d'* was significantly lower in the no-feedback compared to word-feedback condition (1.9 ± .4 vs. 2.7 ± .6; β = -0.81, t = -7.5, p <.001). Similarly, average *d'* in the position-feedback condition (2.0 ± .5) was significantly lower than word-feedback condition (β = -0.69, t = -6.4, p <.001), but not significantly different from the no-feedback condition (β = 0.12, t = 1.1, p =.27). The criterion, however, was not significantly different for any of the comparisons (word-feedback, .9 ± .4, vs. no-feedback, .98 ± .5, β = 0.08, t = .76, p =.45; word-feedback vs. position-feedback, .95 ± .5; β = 0.05, t = 0.47, p =.64; no-feedback vs. position-feedback, β = -0.03, t = 0.11, p =.78).

The overall rate of correction attempts (hit rate) was 19% in the word-feedback, 4% in the no-feedback, and 9% in the position-feedback conditions, respectively. As in Exp 1, average *d'* was significantly lower in the no-feedback compared to word-feedback condition (.43 ± .9 vs. 1.1 ± 1.1; β = -0.82, t = -6.3, p<.001). Similarly, average *d'* in the position-feedback condition (.69 ± .9) was significantly lower than word-feedback condition (β = -0.51, t = -3.9, p<.001). Importantly, the comparison between no-feedback and position-feedback also showed a significant difference between the two (β = 0.31, t = 2.4, p = 0.019). Moreover, in keeping with the results of Exp. 1, and in contrast to the pattern of results for metacognitive judgments, criterion placement changed as a function of condition. Average criterion value was significantly higher in the no-feedback compared to word-feedback condition (2.2 ± .3 vs. 1.8 ± .4; β = 0.39, t = 5.1, p<.001), and in the position-feedback (2.0 ± .3) compared to the word-feedback condition (β = 0.21, t = 2.7, p=.008). The position of the criterion was also significantly higher in the no-feedback compared to the position-feedback condition (β = -0.18, t = -2.4, p = .019).

To summarize, the results of Exp 2 replicated the findings of Exp 1 by showing that the removal of visual feedback made production slower but did not cause a drop in accuracy. There was also a modest drop in error awareness with a similar magnitude to that found in Exp 1 (~20% change from the baseline) and a much steeper decrease in corrections (71% and 79% in Exps 1 and 2, respectively). Patterns of changes to *d'* and criterion as a function of removing visual feedback were identical to those reported in Exp 1. Providing positional information did not —and was not expected to— change error awareness. This information, however, caused a small but significant increase in the rate of corrections compared to when no visual information was present. The criterion in the position-feedback condition also fell in between word-feedback and no-feedback conditions, showing that access to positional information increased participants' motivation to attempt more repairs at the risk of potentially making more False Alarms. In short, positional information, in the absence of any information about letter identity was enough to change participants' repair behavior.

General Discussion

In two experiments, we applied SDT to data from a typing-to-dictation task, to assess the contribution of visual information in general (Exp 1) and positional information in specific (Exp 2) to error awareness and correction. Replicating previous reports (Pinet & Nozari, 2020), we found that removing visual information caused a modest decrease in error awareness and a much stronger decrease in correction rates. In both cases, this was indexed by a lower *d'* when visual information was removed. This finding shows that the overall error signal can be successfully modeled as a combination of internal and external channels, and that the removal of the external channel manifests as increased noise (i.e., closer distributions) in SDT terms.

Participants, however, treated the reduced quality of information differently when making metacognitive judgments about performance vs. when attempting corrections. They only shifted their criterion for a decision in the latter case to avoid False Alarms, because of the cost associated with attempting corrections for an already correct response. This finding shows that although, generally speaking, the same kind of information (broadly defined as conflict) underlies both metacognitive judgments and corrections, (implicit) decisions about how to use such information is task-dependent.

Teasing apart the contribution of positional information from letter identity revealed that positional information alone had a small but significant effect in enhancing corrections; *d'* increased when participants could keep track of where they were in the word without seeing the letters, as in password typing. Inspection of the criterion also showed that having access to positional information increased participants' confidence in aiming for higher Hits, potentially accepting a higher risk of False Alarms.

To summarize, this study demonstrated the utility of SDT in investigating different outcomes of monitoring (error awareness and corrections) in a framework that combined information from internal and external channels, regardless of specific mechanisms. Despite the mechanistic differences postulated in models of language monitoring, our approach allows for drawing general conclusions about the underlying processes, that could then be integrated into, and further investigated within, a specific framework. In particular, this application revealed important commonalities between tasks, i.e., reliance on generally similar information, as well as differences, i.e., different strategies for dealing with the change in information quality. Finally, the framework was useful in hierarchically investigating the finer-grained contributions of specific kinds of information in the visual signal. These results serve two purposes: they shed light on the importance of external information for monitoring performance, especially for applying repairs, and at the same time show the promise of SDT in furthering our understanding of how information from various sources are combined, and how participants handle the partial loss of information in various tasks that depend on monitoring. Given the individual-fitting of the model, this approach is also particularly promising for the analysis of individual differences in monitoring, and monitoring-related functions.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624-652.

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, *73*, 80-94.

Guenther, F.H. (2016). *Neural Control of Speech.* Cambridge, MA: MIT Press.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., ... & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, *48*(3), 829-842.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135-145.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, *82*(13), 1–26.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Makowski, D. (2018). The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. Journal of Open Source Software, 3(22), 470-471.

Nozari, N. (2018). How special is language production? Perspectives from monitoring and control. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of Learning and Motivation* (pp. 179–213). Cambridge, MA: Academic Press.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33.

Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, *36*(5-6), 193-207.

Pinet, S., & Nozari, N. (2018). "Twisting fingers": the case for interactivity in typed language production. *Psychonomic Bulletin & Review*, *25*(4), 1449-1457.

Pinet, S., & Nozari, N. (2020). Electrophysiological correlates of monitoring in typing with and without visual feedback. *Journal of Cognitive Neuroscience*, *32*(4), 603–620.

Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). Typing is writing: linguistic properties modulate typing execution. *Psychonomic Bulletin & Review*, *23*(6), 1898-1906.

Wixted, J. T. (2020). The forgotten history of signal detection *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201–233.