Browser-based Speech Production: Comparing continuous picture naming with spoken and typed response modalities

Kirsten Stark[1,2,3*], Cornelia van Scherpenberg[3,4,5*], Hellmuth Obrig[3,4,5], Rasha Abdel Rahman[1,2,3]

1 Einstein Center for Neurosciences, Charité – Universitätsmedizin Berlin

2 Department of Neurocognitive Psychology, Humboldt-Universität zu Berlin

3 Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Germany

4 Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences

5 Clinic for Cognitive Neurology, University Hospital and Faculty of Medicine Leipzig

* Shared first authorship

First author email addresses:

kirsten.stark@hu-berlin.de scherpenberg@cbs.mpg.de

For experimental research on speech production, temporal precision and high quality of the recorded audio-files are mandatory. These requirements are a considerable challenge if speech production is to be investigated online (Anwyl-Irvine, PsyArXiv, 2020; Bridges, PeerJ, 2020). However, besides the current situation, online research has a huge potential regarding efficiency, ecological validity and diversity of study-populations in psycholinguistic and related research. Here, we supply confirmatory evidence that language production can be investigated online (see also Fairs & Strijkers, PsyArXiv, 2021; Vogt et al., PsyArXiv, 2021). Moreover, we demonstrate that the written naming responses (using the participants' computer keyboard) are a reliable and efficient alternative to typical overt spoken responses in an established speech production paradigm.

To assess semantic interference effects in both modalities we performed two pre-registered experiments (*n*=30 each, sample size estimated based on power analyses) in online settings using the participants' web-browsers and the experimental platform SoSciSurvey (Leiner, 2019; Figure 1A). A cumulative semantic interference (CSI) paradigm was employed that required naming several exemplars of semantic categories embedded in a seemingly unrelated sequence of objects (Figures 1B+C). Previous research (e.g., Howard et al., Cognition, 2016; Costa et al., PNAS, 2009; Schnur, J Mem Lang, 2014) has shown reaction time (RT) to increase linearly for each additional exemplar of a category.

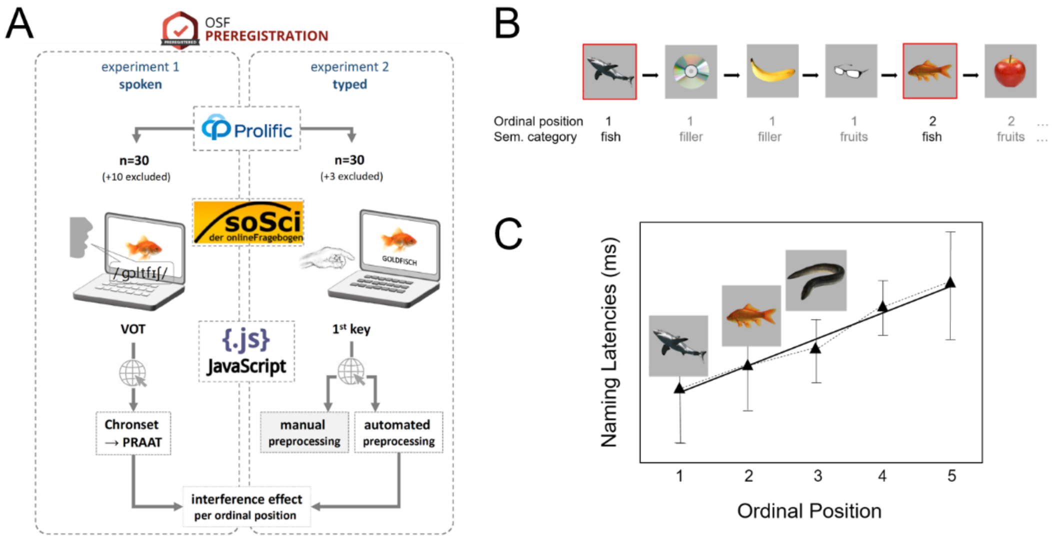
In experiment 1, cumulative semantic interference effects in naming times described in lab-based studies were replicated (Figure 2). In experiment 2, the responses were typed on the participants' computer keyboards and first correct key press was used for RT analysis. While the overall reaction times were slower, this novel response assessment yielded a qualitatively identical, very robust CSI effect. Besides technical ease of application, collecting typewritten responses allows for automatic data preprocessing using string matching procedures and considerably reduces work load for language production research, potentially not only in web-based contexts.

While it has previously been shown that web-based assessment of keyboard responses is reliable for within-participant comparisons (Pinet et al., Behav. Res. Methods, 2017) and research suggests that typed and spoken responses are influenced similarly by linguistic variables (e.g., Pinet et al., Psychon, 2016; Torrance et al., Behav. Res. Methods, 2018), systematic comparisons of both modalities in typical semantic interference paradigms are scarce, let alone their implementation in an online context.

Thus, results of both experiments open new perspectives for research on RT-effects in language experiments across a wide range of contexts in both lab- and browser-based settings. JavaScript- and R-based implementations for data collection and processing are available for download.

Figure 1

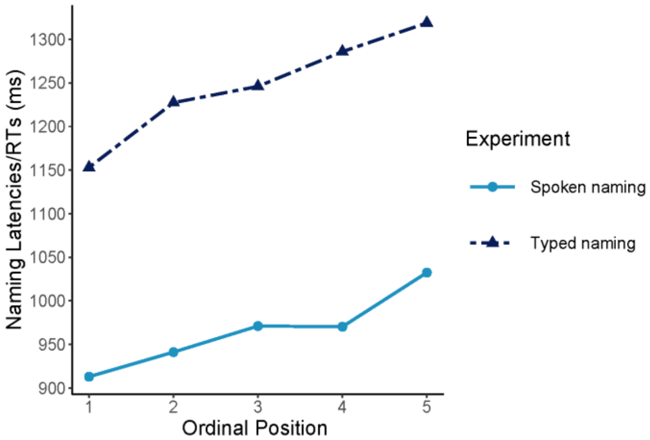
Overview of the Experimental Procedures in Experiments 1 and 2 (A) and Schematic Illustration of the Cumulative Semantic Interference Paradigm (B) and Effect (C)



Note. **A.** The structure of both pre-registered experiments was kept as similar as possible. In both experiments, 30 participants were recruited via the recruitment website Prolific. Participants were then linked to the experimental platform SoSciSurvey where they performed the picture naming task with spoken (experiment 1) or typewritten (experiment 2) responses. Reaction times of voice onset (VOT) and first correct keypresses, respectively, were measured using audio file recordings (experiment 1) and JavaScript (experiments 1+2; Khan, 2020, github.com/muaz-khan/RecordRTC; Stark, 2021, https://github.com/kirstenstark/typing_RTs_JS). After downloading the data from the web server, audio files were preprocessed using Chronset and Praat. For typewritten responses, manual and automatic preprocessing procedures in R (Stark, 2021, github.com/kirstenstark/stringmatch_typed_naming; van der Loo, R J., 2014) were compared and found to be highly correlated (Pearson's r = .97). Finally, the cumulative semantic interference effect on reaction times was analyzed using generalized linear mixed models in R. **B.** In the continuous picture naming task, participants named 160 object pictures from 24 different semantic categories (à 5 exemplars each, plus 40 unrelated fillers). **C.** Based on previous research, RTs are expected to increase linearly with ordinal position, i.e. with each additional member of a semantic category being named. In the example, goldfish and apple would thus be named slower than shark and banana, which in turn would be named faster than the third to-be-named member of a category (e.g., eel, strawberry).

Figure 2

Mean Naming Latencies (RTs) in Milliseconds as a Function of Ordinal Position for Spoken and Typewritten Responses



Note. In both modalities, the expected linear increase of reaction times was replicated. While the overall reaction times in typewritten responses were slower than in spoken responses, the cumulative interference effect (the linear increase in RTs by ordinal position) was qualitatively similar. The interaction between experiment and ordinal position was not significant. Please note that the modality differences in overall reaction times may be confounded with hardware and software effects inherent to browser-based data collection and should therefore be interpreted with care.