Can object difficulty be predicted based on disfluencies and eye movements in connected speech?

Aurélie Pistono\textsuperscript{1}, Robert J. Hartsuiker\textsuperscript{1}

\textsuperscript{1}Department of Experimental Psychology, Ghent University, Belgium

Correspondence

Aurélie Pistono

Ghent University, Department of Experimental Psychology

Henri Dunantlaan 2

9000 Gent (Belgium)

aurelie.pistono@ugent.be
Abstract

In the current study, we asked whether delays in the earliest stages of picture naming elicit disfluency. To address this question, we used a network task, where participants describe the route taken by a marker through visually presented networks of objects. Additionally, given that disfluencies are arguably multifactorial, we combined this task with eye tracking, to be able to disentangle disfluency related to word preparation from other factors (e.g., stalling strategy). We used visual blurring, which hinders visual identification of the items and thereby slows down selection of a lexical concept. We tested the effect of this manipulation on disfluency production and visual attention. Blurriness did not lead to more disfluency on average and viewing times decreased with blurred pictures. However, multivariate pattern analyses revealed that a classifier could predict, from the pattern of disfluency, whether each participant was about to name blurred or control pictures. Impeding the conceptual generation of a message therefore affected the pattern of disfluencies of each participant individually, but this pattern was not consistent from one participant to another. Additionally, some of the disfluency and eye movement variables correlated with individual cognitive differences.

Keywords: disfluency, eye tracking, connected speech, network task

Introduction

The term ‘disfluency’ includes various phenomena such as filled or silent pauses, repeated words, and self-corrections. Despite the high frequency of disfluencies (Fox Tree, 1995), the question remains as to why speakers are so often disfluent. Within the language system, several of the language production levels may be involved in the production of disfluencies but to date, no psycholinguistic model of language production takes into account disfluencies or explains mechanistically why they occur. Influential models of language production (Dell, 1986; Levelt,
1999) simulate speech errors and reaction times, but neither simulates disfluencies. Some studies used a network task (e.g., Hartsuiker & Notebaert, 2010) to investigate where disfluency comes from within the language system. In this task (Figure 2), participants describe a route taken by a point marker through a network of pictures so that a listener could fill in a blank network by listening to the description. This paradigm allows for the manipulation of the items so as to create difficulties at specific production stages (e.g., conceptual generation) while holding other stages constant (e.g., lexical retrieval). It has been shown, for example, that hampering the conceptual generation of a message (Schnadt & Corley, 2006), the verbal monitoring system (Oomen & Postma, 2001) or the initial stage of lexical access (Hartsuiker & Notebaert, 2010) affected the rate of disfluencies. However, not all disfluencies are related to difficulties in speech encoding. For example, some accounts consider disfluencies as a communicative act, used as words to 'signal' delays in speech delivery to an interlocutor (Clark & Fox Tree, 2002). Other authors stressed the role of individual differences in executive function and verbal intelligence on the fluency of speech. In particular, Engelhardt, McMullon, and Corley (2018) found that repetitions were significantly related to individual differences in verbal intelligence, while silent pauses and self-corrections, were marginally related to working memory measures. The current paper asks whether we can tease apart the multiple factors (i.e., conceptual formulation, communicative act, or individual cognitive differences) involved in disfluency production during network descriptions.

One possible way of unravelling the mechanisms underlying disfluency production is the use of eye tracking. Indeed, speakers’ eye movements are revealing about what their language production systems are working on. In particular, the eye-voice span before speech onset (Onset-EVS; the latency from the onset of the first fixation on a picture to the acoustic onset of its name) increases with word preparation difficulty, for example before low name agreement
pictures (Griffin, 2001), pictures with low-frequent names, or visually degraded pictures (Meyer et al., 1998). Additionally, eye movements follow the order of picture description: from 100 to 300 ms before saying an object’s name, speakers shift their gaze to the next object to be named (Griffin, 2004). This lag has been argued to coincide with the end of phonological encoding. Regarding disfluencies more specifically, eye tracking indeed proved to be informative about the mechanisms underlying these phenomena, beyond difficulties related to speech encoding. In a previous study (Pistono & Hartsuiker, 2021), we combined a network task with eye tracking, to evaluate the effects of lexical and grammatical selection on disfluencies and eye movements, during a task that elicits connected speech. We showed that onset-EVS increased with lexical selection difficulty, which suggests that these latencies reflect the time required for word preparation during connected-speech production as well. We also revealed a strong connection between disfluencies and eye movements: some disfluencies were associated with onset-EVS, which implies that they reflect speech encoding difficulties, while others occurred as a stalling strategy. Indeed, participants sometimes inspected other areas than the upcoming picture, while producing pauses. This suggests they made use of “strategic pauses”, consistent with the account of disfluency proposed by Clark and colleagues (e.g., Clark & Fox Tree, 2002). In sum, eye tracking confirmed that not all disfluencies arise from troubles within the language production system. Moreover, (Pistono & Hartsuiker, 2021) also tested whether patterns of disfluency in response to specific difficulties are robust and predictable enough to allow an algorithm to predict which difficulty triggered them. More specifically, we tested whether the manipulated difficulty could be predicted based on the pattern of disfluency associated with it, using multivariate pattern analyses (MVPA, Haynes & Rees, 2006). Instead of analyzing each dependent variable individually, MVPA extracts the information contained in the pattern of information available, to test whether experimental conditions can be distinguished from one another on the basis of the patterns observed. Applying this method to
a network task indicated that lexical selection difficulty could be decoded above chance level from the pattern of disfluency and the pattern of eye movements produced by a speaker. Thus, both patterns of disfluencies and eye movements are sufficiently informative about this linguistic difficulty that a classifier could learn and predict the type of item a speaker was about to mention.

The current study focuses on the conceptual preparation of the speech message, to examine the pattern of disfluencies and eye movements related to a difficulty occurring at this level. It also tests whether individual nonverbal cognitive differences and strategies from the speakers influence the production of disfluencies and eye movements during a connected speech task. To do so, we used the same paradigm as Pistono and Hartsuiker (2021). In order to influence the ease of conceptual preparation of the message, we manipulated visual clarity of the target objects through visual blurring. Regarding the effect of difficulty as the conceptual level, we expected more prolongations before naming blurred pictures, as shown in previous work (Schnadt & Corley, 2006). However, contrary to these authors, we first conducted a preliminary experiment that allowed us to choose the best threshold of blurriness required to elicit difficulty. We also expected blurred pictures to induce longer gaze durations prior to naming in a network task, as shown during picture naming (Meyer et al., 1998, using degraded contours). Furthermore, we tested whether difficulty in conceptual preparation could be predicted based on the pattern of disfluency and eye movements, using MVPA. We had no hypothesis regarding which of these two measures could best predict conceptual preparation difficulty.

Moreover, we expected that not all disfluencies were related to conceptual preparation difficulties. In particular, because blurriness is a salient visual difficulty, we assumed that it would encourage the use of certain strategies while performing the task (i.e., anticipatory
fixations towards blurred pictures to compensate for visual processing difficulty). We also predicted negative correlations between participants’ cognitive performance (i.e., inhibition, working memory, and cognitive flexibility) and the proportion of disfluencies and the time spent on pictures during the task.

Figure 1. Example of a network.

1. Preliminary Experiment

In order to interfere with the conceptual formulation of the message, we aimed to make identification of object images more difficult, by blurring them. We first conducted a preliminary experiment to choose the degree of degradation that was required to slow down naming while retaining high identification accuracy. Indeed, psycholinguistic studies using degraded pictures usually do not provide detailed explanations about the blurring method they used, and the degree of blurring varies considerably among studies. While Schnadt and Corley (2006) used a Gaussian blur with a 1.5 pixel radius for their network tasks, several picture naming experiments used other thresholds (e.g., Gaussian blur with a radius of 6 pixels in Laws & Hunter, 2006; Gaussian blur with a radius of 10 pixels in Brodeur et al., 2010; etc.). In our
preliminary picture naming experiment, we therefore manipulated different thresholds of
blurriness: Gaussian blur with a radius of 2 pixels, 4 pixels, or 6 pixels. We compared accuracy
and reaction times related to these manipulations and to a control condition (no blurriness). We
used a maximum threshold of 6 pixels as Laws and Hunter (2006) found significantly more
errors for blurred than original images in their experiment. On the contrary, we wanted to ensure
that, in the network task, blurred pictures do not induce significantly more errors or “do not
know” responses. The best threshold to tackle disfluency and eye movements related to
conceptual formulation difficulty will therefore be the one leading to longer reaction times in
comparison to control pictures, but similar accuracy.

1.1.  Methods

1.1.1. Participants. Sixteen students of Ghent University, all native speakers of Dutch,
participated. They received 5€ for their participation.

1.1.2. Material. We selected 160 pictures from the Multipic database (Duñabeitia et al.,
2016), controlled for name agreement, Age of Acquisition, and Visual complexity (see Table 1
for more information). We then applied a Gaussian filter on each picture, with a radius blur of
2 pixels, 4 pixels, or 6 pixels (Figure 2), using Photoshop software.

Figure 2. Example of each condition used for the preliminary experiment.
1.1.3. **Procedure.** Using a program written in Psychopy (Peirce, 2007) each participant named a version of all 160 pictures, under four predefined conditions to ensure all pictures were named equally often at each level of blurriness. In each version, 40 pictures were not blurred, 40 pictures had a radius blur of 2 pixels, 40 pictures had a radius blur of 4 pixels, and 40 pictures a radius blur of 6 pixels. The order of pictures was randomized across participants. Participants were instructed to name pictures as soon as they appeared on the screen.

1.1.4. **Data analysis.** We analysed accuracy and reaction times (RTs) associated with each level of blurriness. RTs were measured from onset of the target picture until speech onset. The Chronset algorithm (Roux et al., 2017) was used to automatically extract RTs, in MATLAB. We excluded trials where participants hesitated before providing a name or produced another name before correcting themselves. Then, linear mixed effects models were performed, by means of the lme4 package in R (Bates et al., 2015). Models were based on the “maximal random effects structure” approach (Barr et al., 2013) and then reduced until a further reduction would imply a significant loss in the goodness-of-fit (Matuschek et al., 2017). Pairwise comparisons were performed with the emmeans package (Lenth et al., 2018), with Tukey-adjusted p-values.

1.2. **Results**

1.2.1. **Descriptive.** Mean reaction time was 1189 ms (± 446 ms) for control pictures; 1212 ms (± 447 ms) for 2 pixels radial blur pictures; 1233 ms (± 446 ms) for 4 pixels radial blur pictures; and 1285 ms (± 477 ms) for 6 pixels radial blur pictures. Regarding accuracy, there were, in total, 412 correct answers for control pictures (64.4%); 411 correct answers for 2 pixels radial
blur pictures (64.2%); 366 correct answers for 4 pixels radial blur pictures (57.2%); and 288 for 6 pixels radial blur pictures (45%).

1.2.2. **Accuracy.** A generalized linear mixed effect model tested for effects of blurriness on accuracy. This measure was tested with a random intercept for subjects, and a random slope for blurriness over items. This resulted in a significant effect of blurriness ($\chi^2 (1)=46.81, p<.0001$). Pairwise comparisons of all possible pairs revealed a significant difference between the control condition and the 6 pixels radial blur ($Z = 5.7, p < .0001$), between the 6 pixels radial blur and the 2 pixels radial blur ($Z = 6.34, p < .0001$), and between the 6 pixels radial blur and the 4 pixels radial blur ($Z = 5.26, p < .0001$).

1.2.3. **RTs.** Log-transformed RTs were tested with a random intercept for subjects and items. This resulted in a significant effect of blurriness ($\chi^2 (1)=45.31, p<.0001$). Post-hoc pairwise comparisons of all possible pairs showed a significant difference between the control condition and the 6 pixels radial blur ($t = -6.39, p < .0001$); between the control condition and the 4 pixels radial blur ($t = -3.22, p < .01$). Additionally, the 6 pixels radial blur differed significantly from the 2 pixels radial blur ($t = -5.27, p <.0001$) and the 4 pixels radial blur ($t = -3.39, p < .01$).

1.3. **Discussion**

The analysis of accuracy showed that the 6 pixels radial blur induced significantly more errors than other manipulations, similarly to (Laws & Hunter, 2006). Reaction times were significantly longer for the 4 pixels and 6 pixels radial blur. Based on these findings, we considered the 4 pixels filter to be the best manipulation for the network task. Indeed, word preparation seems more difficult with this filter (i.e., more time required to prepare the answer), but participants were still able to provide the right name, which makes a comparison with a
control condition possible. On the contrary, the 2 pixels radial blur did not vary from the control condition on any of the variables, which implies that this manipulation is not challenging enough.

2. Main Experiment

The main experiment aimed at examining the pattern of disfluencies and eye movements related to conceptualization difficulty (i.e., blurriness). We expected more prolongations and longer gaze durations prior to naming blurred pictures. We further expected that difficulty in conceptual preparation could be predicted based on the pattern of disfluencies and eye movements (using MVPA). Finally, we predicted that individual nonverbal cognitive abilities are correlated with disfluency production and eye movements.

The methods described here are similar to the one used in Pistono and Hartsuiker (2021).

Data, scripts and written transcripts are made available here: https://osf.io/9yhcb/

2.1. Methods

2.1.1. Participants. The samples have been calculated using guidelines for mixed models in designs with repeated measures, for which 1600 observations per condition are required (Brysbaert & Stevens, 2018). For the current experiment, at least 20 participants needed to be included (i.e., 20 participants*((20 networks*8 images)/2 conditions) = 1600 observations per condition). Given that observations had to be excluded if the participant used the wrong name, did not produce anything, or omitted the determiner, we aimed at testing 25 participants to ensure a sufficient power. However, we had to stop the recruitment before we reached this sample size, due to restrictions on lab-based experiments as a result of the Covid-19 pandemic.
Twenty bachelor students, all native speakers of Dutch, participated in the experiment in exchange for course credit (18 Females and 2 Males, mean age was 18.6±0.6 years old).

2.1.2. **Material.** We constructed 20 networks using a program written in Psychopy (Peirce, 2007). We used the 160 pictures from the dataset presented in the preliminary experiment. This dataset was split in two sets, matched for name agreement, age of acquisition, and visual complexity (Table 1). These parts were counterbalanced across participants, so that each part was the control condition for half the participants.

Each network consisted of eight interconnected black-and-white line pictures: four blurred (4 pixels radial blur) and four control pictures (Figure 1). Within the network, pictures were either connected by one, two, or three straight lines or curves. Lines were either horizontal, vertical, or diagonal. Curves could also be horizontal, vertical, or diagonal. The type and number of lines connecting the pictures, as well as the order and location of appearance of the 160 pictures were randomized across participants. The route through the network was indicated by a moving red dot that traversed the network in 42 seconds.

Table 1. Mean (±SD), age of acquisition (AoA), visual complexity, name agreement (H-statistic), in isolation for each set of pictures (i.e., each set was the control condition for half the participants).

<table>
<thead>
<tr>
<th></th>
<th>Set1</th>
<th>Set2</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AoA</td>
<td>6.2±1.1</td>
<td>6.2±1.3</td>
<td>0.96</td>
</tr>
<tr>
<td>Visual complexity</td>
<td>3±0.6</td>
<td>3±0.5</td>
<td>0.62</td>
</tr>
<tr>
<td>Name agreement (H-statistic)</td>
<td>0.8±0.2</td>
<td>0.8±0.2</td>
<td>0.83</td>
</tr>
</tbody>
</table>
2.1.3. **Apparatus.** The experiment was implemented using Psychopy to display networks and record both eye tracking and speech production. Eye movements from participants' dominant eye were monitored with an EyeLink 1000+ system, with a sampling rate of 500 Hz. The monitor display resolution was 1921 x 1081 pixels. Pictures’ resolution was 150 x 150 pixels, subtending 2.8° of visual angle. Head movements were minimized with chin/head rests.

2.1.4. **Procedure.** The participants were tested individually in a quiet room. They took their place in front of a computer screen, which displayed an example network. Instructions were given to provide an accurate description of the network while staying synchronized with the dot that moved through the network. Instructions emphasized that a complete description mentioned the route of the dot, including the shape and direction of the lines or curves, and including the objects. Participants were told that their descriptions would be played to listeners who had to fill in an empty network, which only showed the position of the objects. Subsequently, three practice networks were run. The first network was described by the experimenter to illustrate the task and the next two networks were described by the participant. During this training phase, participants were already using the chin/head rest, to adapt the apparatus for the experiment if needed. During the experiment, each network was preceded by a fixation cross in the upper centre of the computer screen and started with a two seconds period for visual inspection after which the movement of the dot started. Each experiment was split in two runs of ten networks to perform a recalibration halfway through the experiment.

After the experiment, participants underwent several cognitive tests of attention, memory, and inhibition: the Trail Making Test (Reitan, 1958), the Corsi block-tapping test (Kessels et al., 2000), and the Stop-signal paradigm (Verbruggen et al., 2008).
2.1.5. **Scoring and data analysis.** All productions were transcribed and scored by a native Dutch speaker. Another native Dutch speaker checked the transcriptions and scores. Disfluencies were noted for utterances related to paths, but only disfluencies preceding picture names were analysed because we manipulated properties of objects (see Table 2). Five types of disfluency were analysed: repetitions (of a sound, syllable, word, or phrase), filled pauses, silent pauses, prolongations, and self-corrections (substitutions, additions, or deletions). Some phenomena (i.e., repetitions and self-corrections) were grouped into broad categories, to ensure a sufficient amount of data within each category. Silent pauses were subjectively defined by raters, as in Hartsuiker and Notebaert (2010). Example are provided in Table 2. One of the transcribers first independently transcribed and scored all networks. In a subsequent phase, a second transcriber listened to all the productions and checked the transcriptions. They disagreed on 17.8% of trials. Disagreements were solved by a third person.

Table 2. Definitions of each disfluency and example from a translated transcription.

Disfluencies that are taken into account in the analyses (i.e., preceding pictures) are in bold.

<table>
<thead>
<tr>
<th>Disfluency</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-correction</strong></td>
<td></td>
</tr>
<tr>
<td>Substitution [//]</td>
<td>When the speaker stops and resumes with a substitution for a word.</td>
</tr>
<tr>
<td>Addition [///]</td>
<td>When the speaker stops and resumes with the addition of new material.</td>
</tr>
<tr>
<td>Deletion [/////]</td>
<td>When a speaker stops without completing an utterance and resumes with a new utterance.</td>
</tr>
<tr>
<td>Other [////]</td>
<td>When the speaker stops and resumes with a grammatical or lexical error.</td>
</tr>
<tr>
<td>Repetition (r)</td>
<td>Repetitions of sounds, syllables, words or (part) phrases.</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Pause</strong></td>
<td></td>
</tr>
<tr>
<td>Silent pause (.): When the speaker delays the speech stream by being silent.</td>
<td></td>
</tr>
<tr>
<td>Filled pause (h): When the speaker delays the speech stream by inserting a filler (e.g., uh, um).</td>
<td></td>
</tr>
<tr>
<td>Prolongation (p): When the speaker delays the speech stream by prolonging a speech sound.</td>
<td></td>
</tr>
</tbody>
</table>

**Example from a transcription**

The ball starts with **the-[common gender] the-[neuter gender] [/]** piano, that there goes to the left via a straight line to **the (p) letter**, then it goes up via a small arch to the left to **the the (r) pig (.)** then (.). from the pig it goes down via a straight a small straight line [//] to **(p) the uh (h) onion**, from the onion it goes to the lef- the right [/] via a small arch down to **the (.) beard**, from the beard it goes with a small arch to the right up to **the needle (. or sword [////]**, from the **sword** it goes up via a straight line to the gorilla (. and from **(p) the gorilla** it goes up via **hm (h)** a small arch to the right to the **doll (.).**

For the analysis of eye movements, fixation positions were categorized by object Areas of Interest (AoI) corresponding to each picture. Five variables were then considered to test the effect of blurriness on eye movements:

- Onset-EVS: Latency (in ms) between the start of the first gaze at the picture and the onset of its name.
- Offset-EVS: Latency (in ms) between the end of the last gaze at the picture and the offset of its name.
- Number of fixations: Number of fixations occurring within the Onset-EVS to Offset-EVS time window.
- Number of anticipatory fixations: Number of fixations that occur on the picture before the dot traversed it.
- Number of late fixations: Number of fixations that occur on the picture after the dot traversed it.

2.2. Results

2.2.1. Descriptive. Of the total number of pictures (N=3200), 40% were excluded because the wrong target was produced or the gender-marked determiner (‘de’ or ‘het’) was omitted. There was at least one disfluency on 36% of the remaining trials: 7.8% (N=150) pictures included at least one self-correction, 10.15% a silent pause (N=195), 5.5% (N=105) a filled pause, and 10.54% a prolongation (N=202). Repetitions were not analysed because there were only 10 observations in this category. Regarding eye movements, 6.7% of trials included at least one anticipatory fixation and 16.7% at least one late fixation. The proportion of disfluency produced on control vs. blurred pictures is described below (Table 3).

Table 3. Percentage of disfluency produced for control vs. blurred pictures (in proportion of all disfluency)

<table>
<thead>
<tr>
<th></th>
<th>Control condition</th>
<th>Blurred condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-corrections</td>
<td>11.2% (N=74)</td>
<td>11.8% (N=78)</td>
</tr>
<tr>
<td>Silent pauses</td>
<td>15.2% (N=100)</td>
<td>14.3% (N=94)</td>
</tr>
<tr>
<td>Filled pauses</td>
<td>10.2% (N=67)</td>
<td>7.0% (N=46)</td>
</tr>
<tr>
<td>Prolongations</td>
<td>15.0% (N=99)</td>
<td>15.3% (N=101)</td>
</tr>
</tbody>
</table>
2.2.2. **Disfluency.** The effect of blurriness on disfluency (all phenomena together) was tested using linear mixed effects (lme4 package in R). For the random part of the model, the maximal random effects structure (Barr et al., 2013) was included. We then chose a backward-selection heuristic, as in the preliminary experiment (Matuschek et al., 2017). This resulted in a random intercept for subjects, network order, and image order, and a random slope for blurriness over subjects. There was no significant effect of blurriness ($\chi^2 (1) = 1.49$, $p = .2$).

*Per disfluency category.* Generalized linear mixed effects model tested for effects of blurriness for each disfluency type (binomial) separately. There was no significant effect of blurriness when each disfluency type was analysed individually (see Appendix 1).

2.2.3. **Eye movements.** Onset-EVS were log-transformed. This measure was tested with a random intercept for subjects, items, image order, and network order. This resulted in a significant effect of blurriness ($\chi^2 (1) = 7.87$, $p < .01$), indicating longer onset-EVS before control pictures. Offset-EVS, number of fixations and late fixations were tested with a random intercept for items, image order, and network order, and a random slope for blurriness over subjects. The number of fixations decreased with blurriness ($\chi^2 (1) = 3.89$, $p < .05$; control pictures: $9 \pm 3$; blurred pictures: $8 \pm 3$) whereas offset-EVS ($\chi^2 (1) = 0.35$, $p < .6$) and late fixations ($\chi^2 (1) = 0.4$, $p < .5$) did not vary with blurriness. The presence of anticipatory fixations was tested with a random intercept for subjects and items. This variable was not significantly affected by blurriness ($\chi^2 (1) = 2.01$, $p < .2$).

2.2.4. **Excluded trials.** Given the high proportion of excluded trials, the same types of analyses were conducted on these trials in order to test the reliability of previous results. 57.4% of excluded trials were blurred pictures. In this data set, 29.99% of pictures elicited at least one
disfluency. Disfluency (all phenomena) was tested with a random slope for blurriness over subjects and a random intercept for items, image order, and network order. There was no significant effect of blurriness ($\chi^2 (1) = 2.02, p = .16$). There was no significant effect of blurriness either when each phenomenon was analysed individually (see Appendix 1). Regarding eye-movements, Onset-EVS, was tested with a random intercept for subjects, items, image order, and network order. This resulted in a significant effect of blurriness ($\chi^2 (1) = 5.35, p < .05$), confirming the effect of longer onset-EVS before control pictures. The number of fixations was tested with a random intercept for subjects, items, and network order. This variable was not significantly affected by blurriness ($\chi^2 (1) = 0.005, p = .94$), which is the only difference with the included set of data.

2.2.5. **Multivariate pattern analyses of disfluency and eye movements.** To investigate whether the conceptual manipulation could be identified based on the pattern of eye movement or disfluency, we performed multivariate pattern classification, using the Scikit-learn toolbox (Pedregosa et al., 2011). Classifiers were trained for each participant to identify whether she was about to mention a control or blurred item. We trained a linear discriminant analysis (LDA) classifier on four disfluency features (i.e., self-corrections, silent pauses, filled pauses, prolongations) and five eye movement features (i.e., onset-EVS, offset-EVS, number of fixations, anticipatory fixations, late fixations). In all MVPA analyses, features were normalized into Z-scores. The classification was performed in a leave-one-out cross-validation approach to ensure unbiased evaluation of classification performance: In a cross-validation fold, the classifier was trained on data from all but one trial and used on the left-out trial to predict its class membership. This procedure was repeated until each trial’s class has been predicted. Accuracy was the proportion of correctly classified trials. Classification accuracies for each analysis were compared to chance level, that is 50% for a two-class problem, using a
one-tailed t-test. To determine which features played a significant role at a group level, we then performed one-sample t-tests on each feature’s contribution in the classification (Haufe et al., 2014).

**Disfluency.** Classification accuracies were significantly above chance level when analysing disfluencies (53.24% on average; t(18)=2.77, p<.01, Figure 3). However, none of the four features was significant at the group level. This means that, although the pattern of disfluency could predict the type of item a participant was about to mention, this pattern was not consistent from one participant to another.

**Eye movements.** Classification accuracies were also above chance level when analysing eye movements (53.09% on average; t(18)=2.21, p<.05, Figure 3). In particular, onset-EVS was a significant feature for this classification (t(19)=2.6, p<.05). The number of fixations was marginally significant (t(19)=2.05, p=.054).

Figure 3. A) Classification accuracy for each participant. The dashed line indicates the level of chance. B) Contribution of each feature when classifying the pattern of disfluency. C) Contribution of each feature when classifying the pattern of eye movements. The white star indicates significance.
2.2.6. **Correlations with cognitive tests.** For each condition separately, we performed Kendall correlations between cognitive tests and disfluency (i.e., proportion of each phenomenon), and eye movements (i.e., mean onset-EVS and mean number of fixations). We chose four cognitive scores: execution time during TMT A (in seconds), execution time during TMT B-A (in seconds), Backward span, and Stop-signal reaction times (SSRT, in ms).

For the control condition, the production of disfluency was not correlated with cognitive performance. Inhibition (ie., SSRT) was correlated with the mean number of fixations ($r=-0.42; \ p=0.01$, Figure 4), indicating more fixations on control pictures for participants with shorter reaction times and thus better inhibition capacities. Cognitive flexibility (TMT B – A) was negatively correlated with Onset-EVS, indicating more fixations on control pictures for participants who had better cognitive flexibility ($r=-0.34; \ p=0.04$).

For the blurred condition, inhibition (SSRT) was correlated with the proportion of silent pauses ($r=-0.42; \ p=0.011$) and with the mean number of fixations ($r=-0.46; \ p=0.005$, Figure 4), indicating more silent pauses and fixations on blurred pictures for participants who had better inhibition capacities. On the contrary, processing speed/simple attention (TMT A) was correlated with the proportion of prolongations ($r=-0.34; \ p=0.04$), indicating more prolongations on blurred pictures for participants with longer execution time during the TMT part A (see Table 4 for detailed results).
Figure 4. Correlations between SSRT and mean number of fixations for A) the control condition B) the blurred condition

Table 4: Correlations between cognitive performance and disfluency and eye-movements in each condition.
2.2.7. **Correlations with preliminary experiment.** The current findings did not confirm our hypothesis of more disfluencies and longer onset-EVS with blurred pictures. To assess whether these variables are sensitive to picture production difficulty, we performed post-hoc Kendall’s correlations between the mean reaction time associated with each item during the preliminary experiment, and the mean onset-EVS and mean number of fixations, associated with each item during the main experiment. These correlations were performed for the 160 items, in each condition separately.

For blurred pictures, there was a positive correlation between mean reaction times during the preliminary experiment and mean number of fixations during the main experiment, but this was not significant after Bonferroni-Holm corrections ($r=0.12$, $p=0.04$). Regarding control pictures, there was a positive correlation between mean reaction times during the preliminary experiment and mean number of fixations ($r=0.23$, $p<0.001$) during the main experiment. There was also a positive correlation with onset-EVS ($r=0.11$, $p<0.046$), but this was not significant after Bonferroni-Holm corrections. In other words, items that induced longer reaction times during the preliminary experiment also induced longer onset-EVS or more fixations during the main experiment, in both conditions.

2.3. **Discussion**
This study showed that blurred pictures did not elicit more disfluency than control pictures overall during a network task. Additionally, time spent on pictures was shorter for blurred pictures than for normal pictures, contrary to what was expected. However, multivariate pattern analyses were able to predict whether participants were about to name a blurred or non-blurred picture based on their pattern of disfluency or eye movements associated with it. Some of the disfluency and eye movement variables correlated with individual difference measures. In the next three sections, we discuss the results for the disfluencies, eye-movements, and the individual difference measures.

2.3.1. **Disfluency.** Surprisingly, impeding conceptual access of object representations did not elicit more disfluency. However, the rate of disfluency was overall quite substantial (36% of trials had at least one disfluency) compared to studies manipulating lexical selection for example (Hartsuiker & Notebaert, 2010; Pistono & Hartsuiker, 2021). It is therefore possible that, because the manipulation was visually salient, the complexity of the entire task increased, leading to more disfluency in both conditions. Additionally, the considerable proportion of excluded trials (40%) may indicate that participants were less able to monitor their production and repair their errors in this experiment. Indeed, making the identification of half the items more difficult might have hampered the monitoring system (i.e., poor error detection and correction while staying synchronized with the pace of the dot) throughout the whole task, similarly to time pressure (Oomen & Postma, 2001) or divided attention (Oomen & Postma, 2002).

The present results differ from those of Schnadt and Corley (2006), who found an effect of blurriness on prolongations. However, their method was quite different: they chose a different threshold of blurriness (1.5 pixel radial blur), a different set of pictures (Snodgrass &
Vanderwart, 1980), and performed different analyses (ANOVAs). Additionally, Schnadt and Corley’s study was conducted among English speakers. This could contribute to different results, as previous work showed a different pattern of hesitations in English, Dutch, and German (de Leeuw, 2007).

Multivariate pattern analyses yielded complementary findings. They revealed that the classifier could predict, from the pattern of disfluency, whether each participant was about to name blurred or control pictures. This means that disfluencies were sufficiently informative about the linguistic difficulty of an item that a classifier can learn and predict the type of item a speaker is about to name. In other words, impeding the conceptual generation of a message affected the pattern of disfluencies of each participant individually. However, none of the four features (i.e., disfluency type) was affected in a consistent way across participants. This means that, although the pattern of disfluency could predict the type of item a participant was about to mention, this pattern differs from one participant to another. This finding explains why linear mixed models did not reveal significant differences: conceptual difficulty manifests itself differently from one participant to another. Correlations with cognitive performance provided insights into these individual differences, as detailed in section 2.3.3.

2.3.2. **Eye movements.** Contrary to what was expected, the allocation of visual attention did not increase with our blurring manipulation. First, blurred pictures did not induce more anticipatory fixations, contrary to our hypothesis. However, this experiment induced more anticipatory and late fixations in general, compared to our previous experiment that manipulated lexical and grammatical selection difficulty (Pistono & Hartsuiker, 2021). This supports the idea that the blurring manipulation influenced eye movement patterns across the board and not only on the manipulated pictures. Second, control pictures induced longer onset-
EVS and more fixations than blurred pictures. These results differ from Meyer and colleagues (Meyer et al., 1998), who found that degrading pictures significantly affected the mean naming latencies and the mean time spent looking at the objects during a picture naming task. These differences may be due to differences in task and stimuli. Indeed, during real-world scene perception, viewers usually concentrate their fixations on interesting and informative regions, while empty, uniform, and uninformative scene regions are often not fixated (Henderson, 2003). It is possible that a dynamic description task induced findings that are similar to real-world processing, where control pictures required more attention and gaze control than blurred pictures.

Multivariate pattern analyses reinforced findings from linear mixed models: classification accuracies were above chance level when analysing eye movements, and onset-EVS and number of fixations were the most important features for these classifications. The pattern of eye movements was therefore more consistent across participants than the pattern of disfluency, indicating more time spent on control pictures. This result suggests that the differences we observed reveal common underlying cognitive mechanisms that are consistent across participants. The pattern of eye movements, and the number of fixations in particular, has proven to be highly discriminating when classifying different tasks. Using similar multivariate analyses, Kardan and colleagues (Kardan et al., 2015), showed that aesthetic preference induced more fixations than scene memorization, and scene memorization induced more fixations than visual search. The current experiment shows that, within a same task, a classification algorithm can also be successful in predicting the type of information being processed, based on eye movement data.

Taken together, these findings are quite surprising, especially because the preliminary
experiment showed an effect of blurriness on reaction times. That is why we performed correlations between mean items’ RTs during the preliminary experiment and mean items’ onset-EVS and number of fixations. By doing so, we revealed an “item effect”: items that induced longer reaction times during the preliminary experiment tended to induce longer onset-EVS or more fixations during the main experiment, in both conditions. The discrepancy between mean RTs for blurred pictures in the preliminary experiment and mean onset-EVS for these pictures in the experiment shows that the network task is not comparable with a naming task when examining conceptual difficulty. Furthermore, the current results might be influenced by the threshold we used for blurred pictures. Even though this threshold was based on a preliminary picture naming task, a network task involves several differences. During a network task, participants have to build complete sentences while staying synchronized with the pace of the dot, and while processing several pictures displayed on the screen. This task is therefore more complex but more importantly, the size of the pictures displayed on the screen are slightly smaller than ones displayed in the single picture naming task. It is therefore possible that the threshold we used became too difficult in this experimental setting. This could contribute to the high rate of excluded trials, and could explain why current results differ from Schnadt & Corley, 2006 -who used a threshold of 1.5 pixel radial blur- and Meyer et al., 1998 -who used a single picture naming task.

2.3.3. **Correlations with cognitive performance**

Correlations with cognitive performance provided insights into the relationship between disfluency and eye-movements. Indeed, they revealed a positive correlation between the number of fixations and inhibition. The fact that this correlation was significant for both conditions suggests that participants with better inhibition abilities spent more time examining the upcoming picture, regardless of its properties, rather than inspecting other areas of interest.
In other words, participants with better inhibition were better at focusing on the item to be named rather than exploring less relevant items, which is coherent with literature showing the existence of an inhibitory mechanism over visual attentional control (Treisman & Sato, 1990; Wühr & Frings, 2008). However, correlations with disfluency show that not all fixations are due to such mechanisms. Indeed, most importantly, two profiles of participants were found. On one hand, participants with better inhibition produced more fixations and more silent pauses in the blurred condition. On the other hand, participants with slower speed of processing produced more prolongations, while also producing more fixations on blurred pictures. These results suggest that silent pauses and prolongations reflect different mechanisms: while silent pauses reflect better cognitive abilities, prolongations reveal slower processing speed, both leading to an increase in the number of fixations on the picture. The use of silent pauses as marker of better cognitive performance during connected-speech production has previously been shown in different types of narratives (Pistono et al., 2016, 2019). In a sentence repetition task however, the length of silent pauses seems to be correlated with inhibition while the frequency of these phenomena is not (Engelhardt, McMullon, & Corley, 2018). To our knowledge, the production of prolongations has not been investigated in relation with cognitive performance in previous literature, and current results therefore needs to be reinforced with further analyses. However, the current sample size did not allow for predictive analyses such as multiple regressions, and future work should determine more thoroughly the influence of cognitive differences on disfluency production.

2.3.4. Limitations and future directions. This study has several limitations. First, the sample size is rather small given that we had to stop testing participants before we reached the required sample size, and given that many trials were excluded due to bad identification of the object to be named. The sample of participants was also recruited among the same community (i.e.,
bachelor students), which might limit the generalizability of current results. Additionally, other manipulations could be considered to tackle conceptual access. Indeed, blurriness may affect other processes (e.g. visual attention, similarity-based interference, etc) than conceptual formulation. Further work is therefore required to replicate the current findings, and several adjustments can be made to the current paradigm. For example, future work could use pictures that are ambiguous in what concept they denote to impede conceptual access without globally changing the properties of the visual display (i.e., an object shaped like another object). Future studies could also limit the information made available to the speaker (i.e., each picture appearing before the dot traverses it), to limit interference with visual control.

3. Conclusion

Disfluency and eye movements seem therefore quite independent in this experiment, contrary to what has been shown regarding pictures’ name agreement (Pistono & Hartsuiker, 2021). Many studies have shown that the manipulation of variables that affect object naming latencies also affect looking times in a similar way (e.g., lexical frequency, name agreement, etc.). On the contrary, eye movements were not linked with naming difficulties in a network task. Nonetheless, MVPA of disfluency patterns showed that conceptual difficulty manifests itself differently from one participant to another. They therefore point to a need for current models of language production to capture inter-individual variability. Altogether, the current findings open new directions for the use of MVPA to study the language production system.

Acknowledgments

The authors thank the research assistants who transcribed participants’ productions and coded disfluencies. This project has received funding from the European Union’s Horizon 2020
research and innovation program under the Marie Sklodowska-Curie Individual fellowship, grant agreement No 832298.

References


Methods, 162(1–2), 8–13. https://doi.org/10.1016/j.neumeth.2006.11.017


# Appendix 1 - Non-significant results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Random structure</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main analysis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent pauses</td>
<td>random intercept for item, subject, network order</td>
<td>( \chi^2 (1) = 0.67, p=0.41 )</td>
</tr>
<tr>
<td>Filled pauses</td>
<td>random intercept for item and subject</td>
<td>( \chi^2 (1) = 2.6, p=0.11 )</td>
</tr>
<tr>
<td>Self-corrections</td>
<td>random intercept for item and subject</td>
<td>( \chi^2 (1) = 1.35, p=0.24 )</td>
</tr>
<tr>
<td>Prolongations</td>
<td>Random slope for blurriness over items, random intercept for subject, network order and image order</td>
<td>( \chi^2 (1) = 0.12, p=0.73 )</td>
</tr>
<tr>
<td><strong>Excluded trials</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent pauses</td>
<td>random slope for blurriness over subjects, random intercept for item, network order and image order</td>
<td>( \chi^2 (1) = 0.3, p=0.58 )</td>
</tr>
<tr>
<td>Filled pauses</td>
<td>random intercept for item and subject</td>
<td>( \chi^2 (1) = 0.006, p=0.94 )</td>
</tr>
<tr>
<td>Self-corrections</td>
<td>random intercept for item, subject, and network order</td>
<td>( \chi^2 (1) = 2.41, p=0.12 )</td>
</tr>
<tr>
<td>Prolongations</td>
<td>random intercept for subject and item</td>
<td>( \chi^2 (1) = 0.51, p=0.47 )</td>
</tr>
</tbody>
</table>