

Dynamics of inhibitory control during bilingual speech production: An electrophysiological study

Xiaochen Zheng^{a,b,*}, Ardi Roelofs^a, Hasan Erkan^a, Kristin Lemhöfer^a

^a Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

^b International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:

Inhibitory control
N2
Language switching
Switch cost

ABSTRACT

Bilingual speakers have to control their languages to avoid interference, which may be achieved by enhancing the target language and/or inhibiting the nontarget language. Previous research suggests that bilinguals use inhibition (e.g., Jackson et al., 2001), which should be reflected in the N2 component of the event-related potential (ERP) in the EEG. In the current study, we investigated the dynamics of inhibitory control by measuring the N2 during language switching and repetition in bilingual picture naming. Participants had to name pictures in Dutch or English depending on the cue. A run of same-language trials could be short (two or three trials) or long (five or six trials). We assessed whether RTs and N2 changed over the course of same-language runs, and at a switch between languages. Results showed that speakers named pictures more quickly late as compared to early in a run of same-language trials. Moreover, they made a language switch more quickly after a long run than after a short run. This run-length effect was only present in the first language (L1), not in the second language (L2). In ERPs, we observed a widely distributed switch effect in the N2, which was larger after a short run than after a long run. This effect was only present in the L2, not in the L1, although the difference was not significant between languages. In contrast, the N2 was not modulated during a same-language run. Our results suggest that the nontarget language is inhibited at a switch, but not during the repeated use of the target language.

1. Introduction

Bilingual speakers can usually stay in the same language or switch between languages fluently, but this process is not as effortless as it seems. In order to properly speak one language and avoid interference from the other, bilinguals need to control their languages in use. This may be achieved by inhibiting the nontarget language and/or enhancing the target language (Allport and Wylie, 1999; Meuter and Allport, 1999). Nevertheless, it remains unclear how language control unfolds over time. What are the consequences of having used the same language for a prolonged period of time when speakers have to switch to a different language? With repeated use of the same language, speaking becomes easier. This could be a consequence of bottom-up activation of the target language (*language priming*, Grainger and Dijkstra, 1992; Grosjean, 1998, 1999). On the other hand, it is also possible that the level of top-down control adapts with longer use of the same language, making the production of that language easier. In the current study, we investigate the dynamics of language control during language repetition and switching.

Language control is commonly studied using a bilingual picture-

naming paradigm, where speakers are asked to name pictures and switch languages according to a given cue (a flag, a color patch, etc). As concerns naming reaction time (RT), bilingual speakers are usually slower when they have to switch to a different language compared to repeatedly naming in the same language, known as *switch cost*. Paradoxically, switching to the dominant first language (L1) is usually more costly than to the weaker second language (L2) (e.g., Gollan et al., 2014a; Meuter and Allport, 1999). It is hypothesized that during L2 repetition, more top-down control is required to inhibit the dominant L1, or enhancing the weaker L2, compared to L1 repetition. Therefore, it becomes more difficult to overcome the residual inhibition or enhancement when switching back to the L1 than vice versa (Allport and Wylie, 1999; Green, 1998). In a mixed-language context, speaking in the L1 can become more difficult even outside switch trials, and thus become slower compared to the L2 (*reversed dominance effect*; e.g., Christoffels et al., 2007; Costa and Santesteban, 2004; Gollan and Ferreira, 2009; Gollan and Goldrick, 2018; Gollan et al., 2014b; Schotter et al., 2019; Verhoeve et al., 2009; see Declerck and Philipp, 2015, for a more extensive discussion on the relationship between top-down control, asymmetric switch cost, and the reversed dominance effect).

* Corresponding author. Donders Institute for Brain, Cognition and Behaviour, Radboud University, P. O. Box 9104, 6500, HE, Nijmegen, the Netherlands.

E-mail address: x.zheng@donders.ru.nl (X. Zheng).

Inhibition is considered to be one of the main forces of the language control process (e.g., Jackson et al., 2001; Green, 1998; Verhoef et al., 2010). However, evidence seems to diverge on how inhibitory control unfolds over time (Kleinman and Gollan, 2018; Zheng et al., 2018b).

Using the bilingual picture-naming paradigm, Zheng et al. (2018b) observed that top-down control (i.e., inhibition of the nontarget language and/or enhancement of the target language) decreases with repeated use of the same language. In that study, the number of same-language trials before a switch (i.e., run length) was manipulated. Results showed that bilingual speakers' responses were slower and less accurate when switching after a short run compared to a long run. This was explained as follows. With repeated use of the same language, the target language becomes more activated and thus less top-down control is needed. As a consequence, it is harder to switch after a short run (when more control is still applied) compared to a long run (when less control is applied), with more residual control to overcome. Interestingly, the *run-length effect* was only present when switching to the L1 rather than to the L2. The difference between languages seems due to the fact that the weak L2 competes less for selection during L1 repetition and requires less top-down control (i.e., inhibition and/or enhancement). Therefore, when switching back to the L2, less residual control needs to be overcome, regardless of whether the run was short or long. By contrast, producing the L2 requires more control during its repetition, hence more residual enhancement/inhibition needs to be overcome when switching to L1.

A different view on the dynamics of language control has been proposed by Kleinman and Gollan (2018), who argued that inhibition accumulates over time. Using the same picture-naming paradigm, they tracked how naming RTs of the target picture changed as a function of the number of unrelated pictures having been named in the alternative language. Crucially, they considered the increase of RTs *within a mixed-language block* as an index of inhibition, rather than the RTs *within a consecutive same-language run*, as investigated in Zheng et al. (2018b). Their results showed that the more unrelated pictures bilinguals had previously named in the nontarget language, the slower they became in naming pictures in the target language. This global inhibition effect was only found in the L1, but not in the L2. The authors argued that every retrieval in the nondominant L2 hinders subsequent retrieval in the dominant L1, but not vice versa. Interestingly, the run-length effect observed by Zheng et al. was also replicated by Kleinman and Gollan in their study, although its interaction with language was absent.

Does inhibitory control accumulate or decrease over the time course of language switching and repetition? And is it inhibition of the nontarget language, or enhancement of the target language, that drives the run-length effect? Evidence from RTs seems to be limited in this case, as there is no neutral condition to distinguish inhibition from enhancement. New insights can be gathered with the help of EEG, where inhibitory control in language switching is often associated with an N2 effect.

The N2 in event-related potentials (ERPs) is a negative-going component peaking around 200–350 ms after stimulus onset. It is commonly associated with response inhibition, such as withholding the button press in a go/no-go task (Falkenstein et al., 1999; Jodo and Kayama, 1992). With respect to language switching, a larger N2 has been observed for switch trials compared to repeat trials, with the switch-cost effect only present in the L2 (Jackson et al., 2001). This N2 switch effect has a fronto-central scalp distribution which is similar to the no-go N2 (but see Christoffels et al., 2007, for a report where a larger N2 was found on repeat trials than on switch trials, particularly in L1). Jackson et al. (2001) argued that the N2 effect on switching reflects inhibition of the competing nontarget lexicon; greater inhibition of L1 is required when switching to L2 compared to switching to L1. Interestingly, in the same study, a larger switch cost in RT was observed in the L1 rather than in the L2. To explain the difference between the RT and the ERP results, the authors argued that “the frontal N2 reflects processes that are in operation to bring about switching whereas the RT

data reflect the net result of having switched” (Jackson et al., 2001, p. 177). A later study replicated the N2 effect for switch costs (to L2) in bilingual picture-naming (Verhoef et al., 2010), but the reported N2 had a more posterior rather than anterior scalp distribution (see also Folstein and Van Petten, 2008, for a review of the dissociation between the anterior and posterior N2 in the nonlinguistic literature). This posterior N2 effect was interpreted as to reflect the disengagement from the nontarget language: Switching to the L2 requires disengagement of the stronger L1 (therefore a larger N2 effect), which is not the case for switching to the L1 (therefore a smaller or no N2 effect). A similar posterior N2 effect has been reported in monolingual task switching as well (Sikora et al., 2016). In that study, speakers were asked to switch between describing black-and-white pictures with short phrases (“the fork”) and describing colored pictures with long phrases (“the green fork”). Because the short phrases need to be inhibited during the production of long phrases, it was more difficult to overcome such inhibition when participants had to switch back to the short phrases compared to the reverse situation. Therefore, a larger posterior N2 effect was observed during switches to short compared to long phrases.

The interplay between the anterior and posterior N2s may shed light on the question how inhibitory control develops during repeated use of the same language and how such inhibition is overcome during language switching. To this end, we employed the bilingual picture-naming paradigm used in Zheng et al. (2018b) and measured bilingual speakers' EEG during naming. To examine whether inhibitory control accumulates or decreases during language repetition, we compared responses on early vs. late ordinal positions within a same-language run.¹ If the target language gets increasingly activated throughout repetition, we should expect faster responses on late than early ordinal positions. Furthermore, if inhibitory control decreases due to bottom-up priming of the target language, then a decrease in the N2 amplitude, as an index of inhibition, should be observed in late compared to early ordinal positions as well. We expected this N2 effect, if observed, to have an anterior scalp distribution, which is associated with response inhibition (Falkenstein et al., 1999) or language inhibition (Jackson et al., 2001).

It is worth noting that although in the current study we consider the N2 effect to be an index of inhibitory control, the interpretations of the N2 in the language literature are not fully consistent (e.g., Jackson et al., 2001; Sikora et al., 2016; Verhoef et al., 2010). In addition, the N2 has also been interpreted to reflect conflict monitoring (e.g., Christoffels et al., 2007; Morales et al., 2013; see also Nieuwenhuis et al., 2003 for nonlinguistic findings). We will discuss these accounts in more detail in the Discussion.

Besides the investigation of repeat trials, we also looked at switch trials to answer the question whether the run-length effect (i.e., switching is more costly following a short compared to a long same-language run) is due to overcoming inhibition at the switch. To this end, we compared the N2 at switch trials following short vs. long same-language runs. If the process of overcoming inhibition dominates during switching, we would expect to replicate Zheng et al. (2018b) in the behavioral results by finding a larger switch cost in RTs when switching after a short run compared to a long run. Furthermore, a larger N2 should be observed at switches following a short than a long run. The scalp distribution of the N2 switch effect may be either more anterior, reflecting inhibition (Jackson et al., 2001), or more posterior, reflecting disengagement or overcoming inhibition (Sikora et al., 2016; Verhoef et al., 2010).

¹ Note that in the current study, “ordinal position” refers to the position of the trial within a same-language run, which is different from the same term defined in Meuter and Allport (1999).

2. Method

2.1. Participants

Thirty participants took part in the study for course credit or vouchers. All of them were native Dutch speakers, raised monolingually, who spoke English as their most proficient nonnative language. All participants were right-handed and had normal or corrected-to-normal vision. Participants were recruited online using the Radboud research participation system and received study credits or vouchers for compensation. The study was conducted in accordance with the Declaration of Helsinki, was approved by the local ethics committee (Faculty Ethics Committee, Radboud University, ECSW 2015-2311-349), and all subjects provided written informed consent.

Four participants' data were excluded from the EEG analysis due to excessive artifacts, and one additional participant was excluded due to a technical problem during recording. To be consistent, we also excluded their data from the behavioral analysis. This resulted in a final set of 25 participants (seven males).

Table 1 summarizes the language background of the 25 participants as assessed by a questionnaire, and their English vocabulary size measured by the LexTALE test (Lemhöfer and Broersma, 2012).

2.2. Materials

Forty black-and-white line drawings, which represented 40 translation pairs of Dutch–English noncognate words (e.g., the Dutch word “boom” and its English translation “tree”), were used as experimental pictures (see Appendix A). All the pictures were taken from the international picture naming project (IPNP) database (Bates et al., 2003). Based on a pilot study on naming agreement, we replaced two of them with drawings sketched by the first author. Pictures were selected with high naming agreement in both Dutch and English (Bates et al., 2003; Severens et al., 2005) and with high frequency names (CELEX database; Baayen et al., 1995) as selection criteria (log frequency: $Mean_{NL} = 1.54$, $SD_{NL} = 0.55$, $range_{NL} = 0.47–2.91$, $Mean_{EN} = 1.01$, $SD_{EN} = 0.61$, $range_{EN} = 0–2.56$; Naming agreement (H-statistics): $Mean_{NL} = 0.50$, $SD_{NL} = 0.47$, $range_{NL} = 0–1.69$, $Mean_{EN} = 0.31$, $SD_{EN} = 0.40$, $range_{EN} = 0–1.56$). We matched all the Dutch and English picture names as closely as possible on number of syllables ($p = .813$) and phonological onset category, so that possible differences in RTs could not be explained by word length or differences in voice-key sensitivity (e.g., /f/ has a delayed voice-key onset compared to /a/). All the pictures were scaled to 300×300 pixels.

Table 1
Participants' language background and English proficiency.

Characteristic	Mean	SD	Range
Age	22.1	2.7	19–27
Age of acquiring English	9.3	1.8	6–11
Self-rated frequency of using English ^a			
speaking	3.6	1.2	1–5
listening	4.5	0.7	3–5
reading	3.6	1.5	1–5
writing	3.1	1.5	1–5
Self-rated frequency of switching languages ^a			
speaking	2.2	1.1	1–4
Self-rated proficiency in English ^a			
speaking	4.2	0.6	3–5
listening	4.6	0.5	4–5
writing	4.1	0.7	3–5
reading	4.6	0.6	3–5
English vocabulary size			
LexTALE test	81.0	12.2	56–98

Note. SD = Standard Deviation.

^a Self-ratings were given on a scale from 1 = *very rarely/bad* to 5 = *very often/good*.

2.3. Design

There were two types of trials: *switch trials*, where the response language was different from that in the previous trial, and *repeat trials*, where the response language stayed the same. On repeat trials, we compared early vs. late ordinal positions within a same-language run. In the current study, a same-language run had a maximum of six trials. Therefore, we coded trials with the ordinal position 2 and 3 within a run as *early* position, and trials 4, 5, 6 were coded as *late* position (position 1 is a switch). Consequently, 58.33% of the repeat trials were classified as in an early position, and 41.67% as in a late position. On switch trials, we compared naming when switching after short vs. long same-language runs. The run length could be short (i.e., two or three repeat trials before a switch) or long (i.e., five or six repeat trials). Each type of run length occurred an equal number of times. Overall, 23.75% of trials in the experiment were switch trials. A schematic diagram of the experimental paradigm can be found in Fig. 1.

Each experimental list had eight blocks of 80 trials each, in total 640 trials. Each stimulus appeared twice in a block, once in Dutch and once in English. We tried to make sure that each stimulus occurred equally often on a switch trial in both languages and after all run lengths.² The pseudo-randomization of repeat trials was done in each block using the program MIX (van Casteren and Davis, 2006), with the following restrictions: (1) subsequent trials were semantically and phonologically unrelated; (2) repetition of a picture was separated by at least four intervening trials. We also made sure that each item occurred at least once in each type of ordinal positions. A second list was constructed by reversing the block order of the first list.

2.4. Procedure

Participants were seated in a sound-proof booth and the experiment was run using the software *Presentation* (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The background color of the computer screen (Benq XL2420Z, 24-inch screen) was set to grey, with a resolution of 1920×1080 pixels, at a refresh rate of 120 Hz. We first familiarized participants with all the pictures. They were asked to name each picture once in Dutch (block 1) and once in English (block 2); if they were unable to name it, they were told the correct answer and asked to remember it and name it again (block 3). Then we followed with a practice block to familiarize participants with the language cues. The cues were presented as a 100-pixel-wide frame around the picture whose color represented the response language (i.e., red and yellow for Dutch, and green and blue for English, or vice versa). Two colors were used to cue each language to avoid a confound of color switch in the stimulus and language switch in the required response (Heikoop et al., 2016; Lavric et al., 2018; Mayr and Kliegl, 2003; Verhoef et al., 2010). The color of the cue switched in every trial. We counterbalanced the assignment of colors to the response language across participants. The practice block ran for a minimum of 40 trials and stopped when participants' accuracy reached 90%. For the first 20 trials, participants received feedback for the correct response after each trial.

EEG was recorded during the eight experimental blocks of the main experiment. Each trial started with a fixation cross (250 ms), followed by a jittered blank screen (250–500 ms). The picture then appeared in the center of the screen together with the color cue, waiting for a response which would be registered by a voice key (Shure SM-57 microphone). After a valid response or no response within a time limit (2000 ms), the stimulus stayed on the screen for another 550 ms. The next trial began after another jittered blank screen (250–500 ms). We instructed the

² There were in total 152 switch trials within a list. Therefore, eight out of the 40 stimuli ended up occurring three times instead of four on the switch trials, leaving out each run length once in each language (see Appendix A for more details).

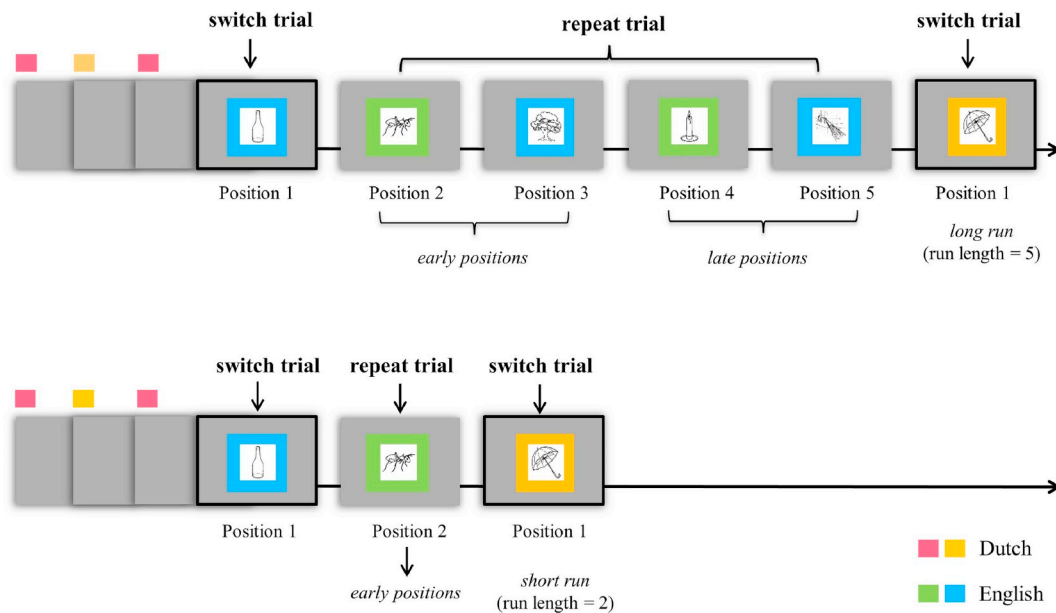


Fig. 1. Experimental paradigm. The length of a run of same-language trials before a switch is short (run length = 2 or 3) or long (run length = 5 or 6). Trials within a same-language run are categorized as early (ordinal position = 2 or 3) or late (ordinal position = 4, 5, 6).

participants to name the pictures as quickly as possible in the language indicated by the cue, and also not to correct themselves when they said something wrong. All the instructions were given in Dutch.

After the main experimental part with EEG measurement, participants completed the LexTALE vocabulary test in English and a language background questionnaire. The entire session took approximately 2 h.

2.5. EEG recording

We recorded EEG from 57 active Ag–AgCl electrodes mounted in an elastic cap, placed according to the international 10–20 system (ActiCAP 64Ch Standard-2, Brain Products). EEG signals were referenced online to the left mastoid electrode and re-referenced offline to the average of the right and left mastoid electrodes. EOG was measured horizontally with two additional electrodes placed above and below the right eye, and vertically with two electrodes placed on the left and right temples. EMG was measured with two electrodes placed next to the upper lip and the throat.³ EEG, EOG and EMG signals were amplified with BrainAmps DC amplifiers (500 Hz sampling, 0.016–125 Hz band-pass). Impedances for EEG electrodes were kept below 20 k Ω .

2.6. EEG preprocessing

We performed all EEG analyses using the Fieldtrip open source Matlab toolbox (Oostenveld et al., 2011) and custom analysis scripts in Matlab v.8.6.0 (R2015b, The Math Works, Inc). We first segmented the continuous EEG into epochs from 200 ms before to 2500 ms after picture onset.⁴ The data were then re-referenced and band-pass filtered with a low cut-off of 0.1 Hz and a high cut-off of 30 Hz. Trials with atypical artifacts (e.g., jumps and drifts) were rejected by visual inspection; EOG artifacts (eye blinks and saccades) were removed using independent component analysis. After that, we further segmented the data into shorter epochs from 200 ms pre- to 500 ms post-picture onset and applied another round of visual inspection to remove trials with

³ EMG was measured to track the time course of speech and to monitor for speech artifacts, but was not analyzed in the current study.

⁴ A long segment was chosen to provide the baseline for response-locked analysis, which was not used in the current study.

remaining artifacts (e.g., muscle artifacts due to early articulation). Baseline correction was applied based on the average EEG activity in the 200 ms interval before picture onset. Individual EEG channels with bad signals were interpolated by a weighted average of the data from neighboring channels of the same participant. On average, we discarded 3.5% of the epochs and 1.5% of the channels. Two channels (FT7, TP7) that were interpolated in more than two participants were excluded from the group-level analyses. We averaged all the epochs for each condition and each participant. Four participants with less than 20 remaining trials in any condition were excluded from the EEG analysis (see also “Participants”).

2.7. Statistical analyses

Participants’ responses were categorized as errors when they used nontarget words (from either language), or when they failed to respond or respond with a repair or disfluency. Errors were excluded from the subsequent RT and ERP analyses. Naming RTs were recorded online using a voicekey and later manually corrected if necessary, using the speech analysis program Praat (Boersma and Weenink, 2016). Correct responses with a RT deviating more than three standard deviations from the respective participants’ condition mean were excluded (per language and per trial type). Trials in the beginning of each block and post-error trials were also excluded. Because a language selection error on a repeat trial (e.g., saying the Dutch word “boom” instead of the English word “tree”) alters the characteristics of the run in which it occurs (e.g., turns a long run into a short run), we decided to exclude all runs with errors. This led to an exclusion of 9.9% of the data. We did not analyze the errors themselves due to (1) their infrequent occurrence (3.9% on the switch trials, 1.8% on the repeat trials, before excluding all runs with errors), and (2) after the exclusion of erroneous runs (see above), there were no error data available for the analysis of repeat trials, nor for the analysis of switch costs.

The statistical analyses of the behavioral data were performed with generalized mixed-effects models using the lme4 package (Version 1.1.13, Bates et al., 2015) in R (Version 3.4.1; R Core Team, 2017) to account for the right-skewed shape of the RT distribution without the need to transform and standardize the raw data (Lo and Andrews, 2015). We started with a full model of RTs as a function of language (L1 vs. L2), trial type (switch vs. repeat), and run length (short vs. long)/ordinal

position (early vs. late), and followed up all the interactions with trial type. To further test our hypotheses on ordinal positions and run length, we also analyzed the RTs of the repeat and the switch trials separately, with language and ordinal positions/run length as factors. For all the analyses, the factors language, ordinal position/run length, and trial type (if applicable) were sum-coded and included as fixed effects. Participants and items were included as random effects. We ran all the models with a maximal random-effects structures, which included random intercepts and random slopes for all fixed effects and their interactions for both participants and items (Barr et al., 2013). Only when the model with the maximal random-effects structure did not converge, we simplified it by first removing the interactions and if necessary the main effects in the random structure.

The statistical analysis of the ERP data was run using a nonparametric cluster-based permutation test (Maris and Oostenveld, 2007) using Matlab v.8.6.0 (R2015b, The Math Works, Inc). The method controls for the false alarm rate caused by multiple comparisons, i.e., when evaluating the ERP data at multiple channels and multiple time points. On the repeat trials, we compared early vs. late positions within a

same-language run; on the switch trials, we compared the switch after short vs. long runs. We also compared the switch cost (i.e., difference between switch and repeat) between languages and between run lengths. To calculate the switch cost between run lengths, early vs. late position of repeat trials were used as a baseline for short vs. long runs of switch trials, respectively.

For the cluster-based permutation test, the two conditions of interest were first compared using a paired-samples *t*-test (two-tailed) at each spatiotemporal sample (i.e., per channel and time point). Then we used an alpha threshold of .05 and all samples with smaller *p*-values are selected. Afterwards, those selected samples which were spatiotemporally adjacent were grouped as clusters. For each cluster, the sum of the *t*-values of all the samples was used as the cluster-level statistic. Using the same procedure as described above, we constructed a permutation distribution by randomly partitioning the original data for 1000 times and then computing spatiotemporal clusters with their cluster-level statistic. We selected the cluster with the maximum cluster-level statistic to compare against the permutation distribution. The *p*-value of the cluster was calculated as the proportion of random partitions (out of

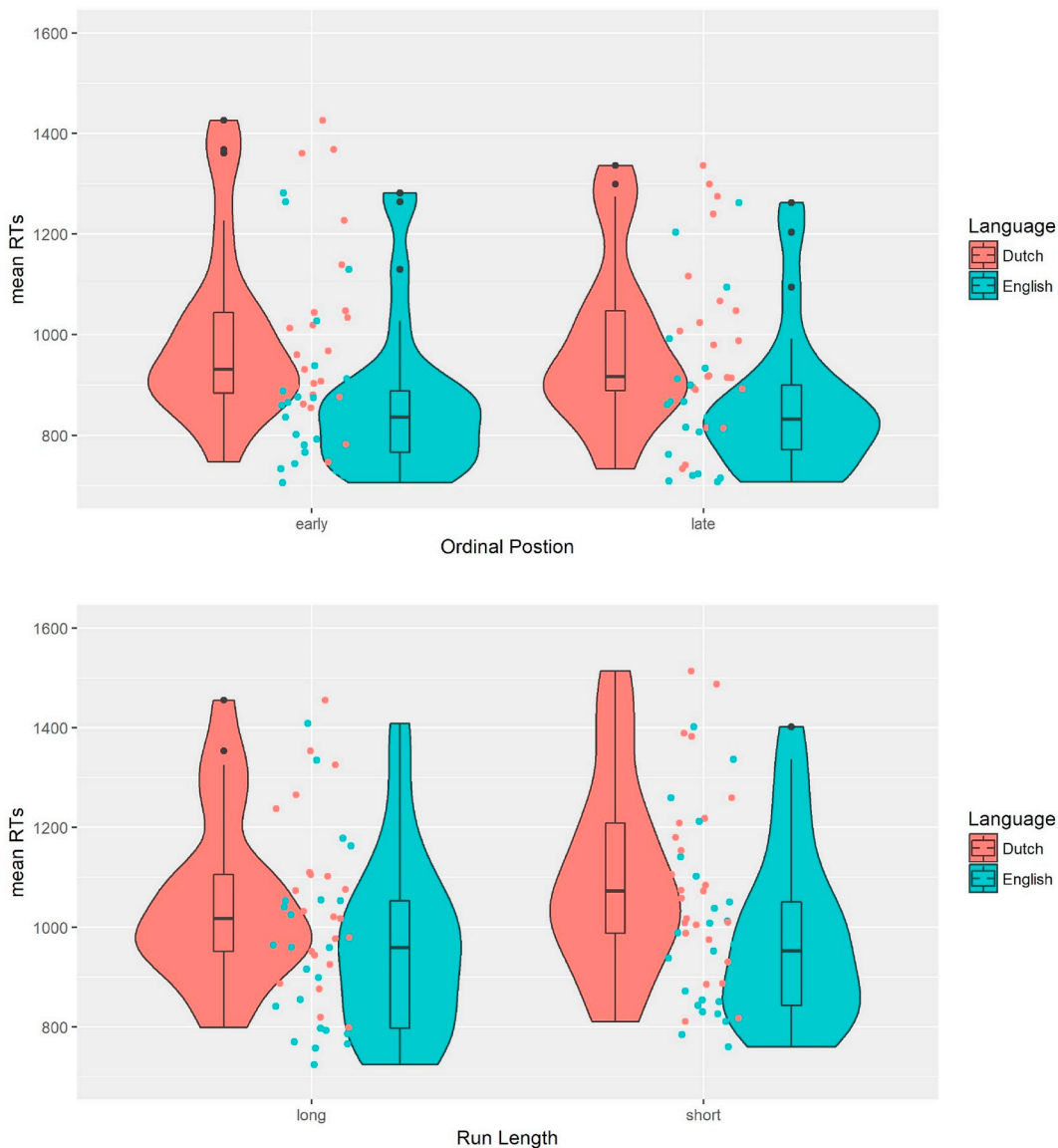


Fig. 2. Violin plots with individual data distributions of mean RTs (in ms) for repeat trials (top panel) and switch trials (bottom panel), grouped by language (Dutch vs. English) and ordinal position (early vs. late, top panel) or run length (long vs. short, bottom panel). The outer shapes represent the distribution of individual data, the thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition.

1000) that yielded a larger cluster-level statistic than its own statistic. A p -value below .05 (two-tailed) was considered to be significant.

We focused our ERP analysis on the N2 components. Following the N2 literature, statistical tests were applied to the time window of 200–350 ms post stimulus onset. Given the two possible topographies of the N2 components (i.e., anterior N2 and posterior N2), we applied our analysis to all available electrodes.

3. Results

3.1. Behavioral results

3.1.1. Overall analysis

Fig. 2 shows the violin plots for the RTs on the repeat trials (top panel) and on the switch trials (bottom panel).

We started with a full model of RT as a function of language (L1 vs. L2), trial type (switch vs. repeat), and run length/ordinal position. For the run length/ordinal position analysis, early vs. late position of repeat trials were used as a baseline for short vs. long runs of switch trials, respectively. Table 2 presents all the statistics from the GLMEM used for this analysis.

Bilingual speakers were slower to respond in L1 than in L2, and on switch than on repeat trials. There was also a main effect of run length/ordinal position: Speakers were faster in late positions/when switching after a long run compared to early positions/when switching after a short run.

Switch costs. We observed an interaction between trial type and language. Follow-up analyses showed that the switch cost (i.e., the difference between trial types) was smaller in L1 ($diff = 87$ ms; $\beta = -41.68$, $SE = 5.06$, $t = -8.23$, $p < .001$) than in L2 ($diff = 105$ ms; $\beta = -55.08$, $SE = 4.36$, $t = -12.62$, $p < .001$). Another interaction was found between trial type and run length/ordinal position: the switch cost was larger for a short run ($diff = 107$ ms; $\beta = -56.80$, $SE = 5.76$, $t = -9.86$, $p < .001$) than a long run ($diff = 85$ ms; $\beta = -43.98$, $SE = 4.42$, $t = -9.96$, $p < .001$).

To test our hypotheses on ordinal positions and run length, we further analyzed RTs separately for repeat trials and switch trials.

3.1.2. Analysis of repeat trials

On the repeat trials, we analyzed how naming RTs differed on early vs. late ordinal positions within a same-language run, and how this interacted with language. The statistics from the GLMEMs for the RTs on the repeat trials are presented in Table 3 (upper half).

On repeat trials, speakers were slower in the L1 than the L2, and in early than late positions. There was no interaction between language and ordinal position ($M_{L1early} = 998$ ms, $SD_{L1early} = 179$ ms; $M_{L1late} = 978$ ms, $SD_{L1late} = 166$ ms; $M_{L2early} = 870$ ms, $SD_{L2early} = 156$ ms; $M_{L2late} = 861$ ms, $SD_{L2late} = 145$ ms).

3.1.3. Analysis of switch trials

On switch trials, we analyzed how naming RTs differed in switches after short vs. long same-language runs, and how this interacted with language. Table 3 (lower half) gives the statistics from the GLMEMs for the RTs on the switch trials.

Speakers were slower to switch after a short run than a long run, and when switching to the L1 (Dutch) than to the L2 (English). There was a significant interaction between language and run length: The run-length effect was present in the L1 ($M_{L1short} = 1101$ ms, $SD_{L1short} = 193$ ms; $M_{L1long} = 1049$ ms, $SD_{L1long} = 167$ ms; $\beta = -28.27$, $SE = 8.28$, $t = -3.42$, $p < .001$), but not in the L2 ($M_{L2short} = 981$ ms, $SD_{L2short} = 177$ ms; $M_{L2long} = 961$ ms, $SD_{L2long} = 177$ ms; $\beta = -10.94$, $SE = 8.79$, $t = -1.24$, $p = .213$).

3.1.4. Summary

Speakers were slower in the L1 than in the L2 – replicating the reversed dominance effect – and when switching compared to a

Table 2

Statistics from the GLMEM for the reaction time (RT, ms) as a function of language (Lang), trial type (Type), and run length/ordinal position (RL/OP).

		Mean (SD)	β	SE	t -value	p
Lang	L1	1032 (180)	60.22	2.50	24.14	<.001
	L2	919 (171)				
RL/OP	short/early	988 (193)	23.78	2.04	11.67	<.001
	long/late	962 (175)				
Type	switch	1023 (185)	-43.21	2.58	-16.75	<.001
	repeat	927 (171)				
Type \times Lang			4.08	1.59	2.57	.010
Type \times RL/OP			-10.84	1.79	-6.08	<.001
Lang \times RL/OP			3.55	1.80	1.98	.048
Type \times Lang \times RL/OP			-1.447	1.755	-0.824	.410

Table 3

Statistics from the GLMEMs for the reaction time (RT, ms) on repeat and switch trials.

			Mean (SD)	β	SE	t -value	p
Repeat	Lang	L1	988 (171)	67.80	2.47	27.44	<.001
		L2	866 (149)				
	OP	early	934 (178)	5.82	2.11	2.76	.006
		late	920 (165)				
Lang \times OP			0.67	1.59	0.42	.672	
Switch	Lang	L1	1075 (180)	54.33	5.98	9.08	<.001
		L2	971 (176)				
	RL	short	1041 (193)	-20.58	5.52	-3.73	<.001
		long	1005 (176)				
	Lang \times RL			-9.57	3.76	-2.54	.001

Note. RL = run length, OP = ordinal position, Lang = language.

repetition of language. Interestingly, switch costs were smaller in L1 than in L2, which seems to contradict previous literature (e.g., Meuter and Allport, 1999). We will address this in the Discussion. On repeat trials, speakers were faster in later than early ordinal positions, suggesting bottom-up activation/priming of the target language. On switch trials, speakers were faster to switch after a long run than a short run. The run-length effect was only present in the L1, not in the L2, replicating Zheng et al. (2018b).

3.2. ERP results

3.2.1. Analysis of repeat trials

Fig. 3 shows the averaged ERPs and topographies for early vs. late ordinal positions within a same-language run, in three representative midline electrodes: Fz (anterior), Cz (central), and Pz (posterior).

The cluster-based permutation tests showed no differential N2 in early compared to late ordinal positions ($p = .378$; Fig. 3A). When further tested within each language (Fig. 3B), no N2 effect was observed for early compared to late ordinal positions either in L1, Dutch ($p = .338$) or in L2, English ($p = .308$). The difference between languages was also not significant, as no clusters were detected in the permutation test.

3.2.2. Analysis of switch trials

Fig. 4 shows the averaged ERPs and topographies for switch trials following short vs. long same-language runs.

The cluster-based permutation tests revealed a significant increase in N2 amplitude on switch trials following a short compared to a long run ($p = .010$, Fig. 4A). The effect was most pronounced between 260 and 350 ms post stimulus onset, with a widespread scalp distribution somewhat centered towards the fronto-central sites and right lateralized. We further compared the N2 effect between the two switching

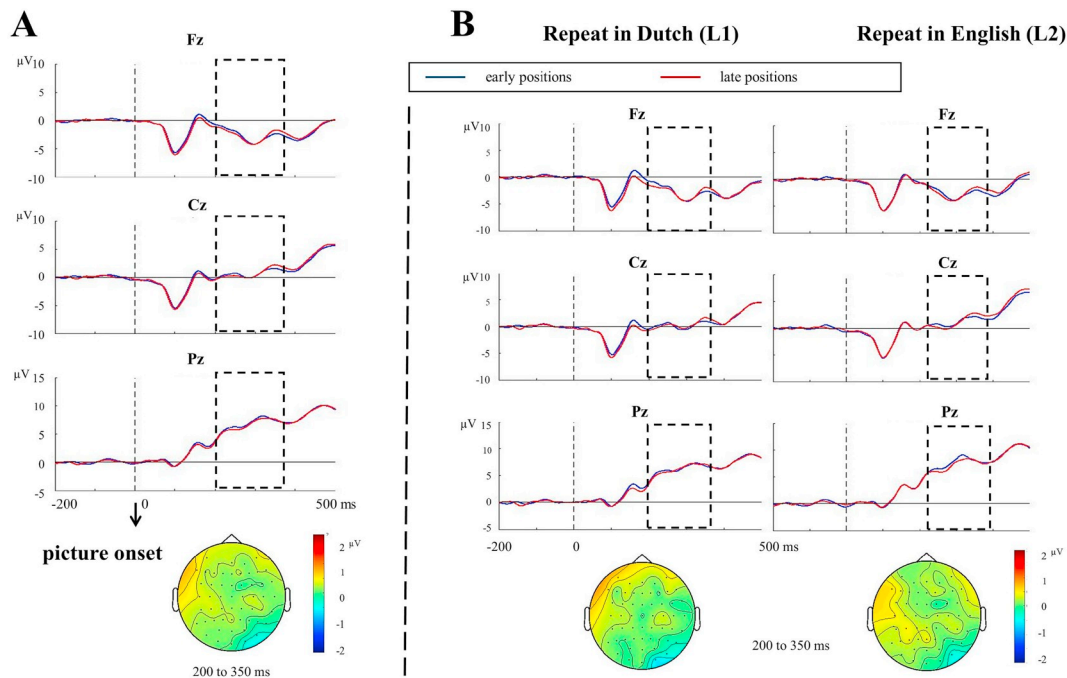


Fig. 3. (A) Stimulus-locked ERPs and topographies for early vs. late ordinal positions, averaged across the two languages. (B) Stimulus-locked ERPs and topographies for early vs. late ordinal positions when repeating in L1 (Dutch) and repeating in L2 (English). The time window used for testing the N2 effect (200–350 ms) is marked by an empty frame. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

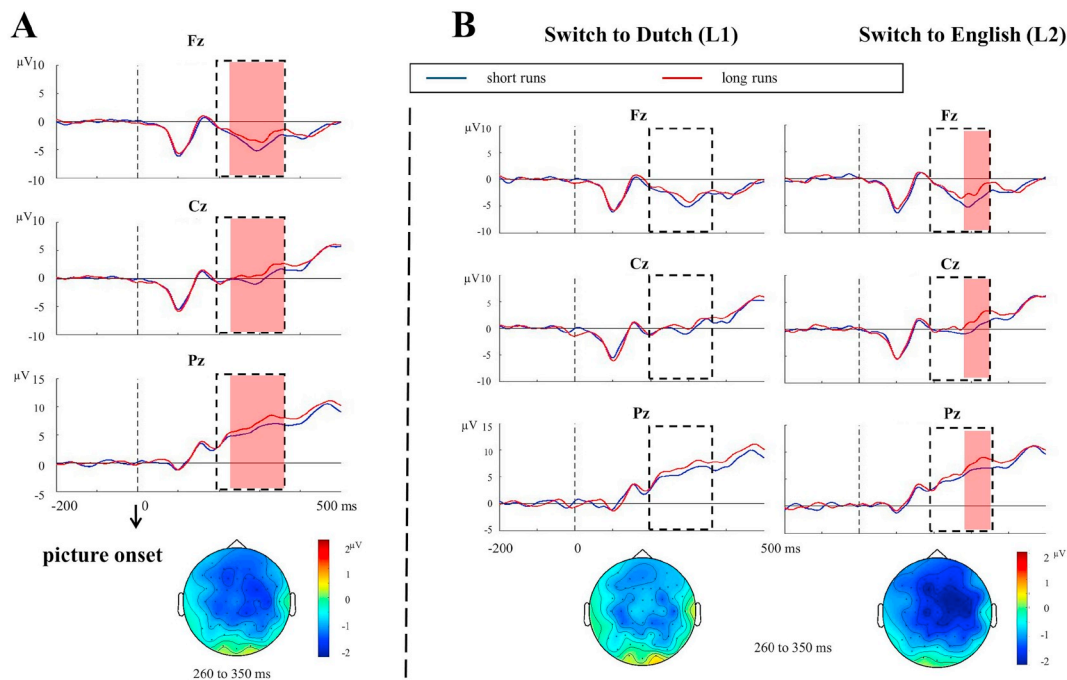


Fig. 4. (A) Stimulus-locked ERPs and topographies for switch trials after short vs. long runs, averaged across languages. (B) Stimulus-locked ERPs and topographies for switching to L1 (Dutch) and to L2 (English) after short vs. long runs. The time window used for testing the N2 effect (200–350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. We used the same time window in A and B (the window that was associated with the N2 effect in A) for depicting the topography for the sake of better comparability.

directions (Fig. 4B). Results showed that the N2 run-length effect was only present when switching to the L2 ($p = .042$). The effect was most pronounced between 320 and 350 ms post stimulus onset, widely distributed over the scalp, and slightly stronger in fronto-central sites and right lateralized. In contrast, the N2 run-length effect was not observed when switching to the L1 ($p = .178$). However, this difference

between languages was not significant ($p = .360$).

3.2.3. Analysis of switch costs

To better compare the current ERP results with previous literature (e.g., Verhoef et al., 2010), we contrasted the switch trials with the repeat trials (i.e., switch costs), as well as the switch costs between languages

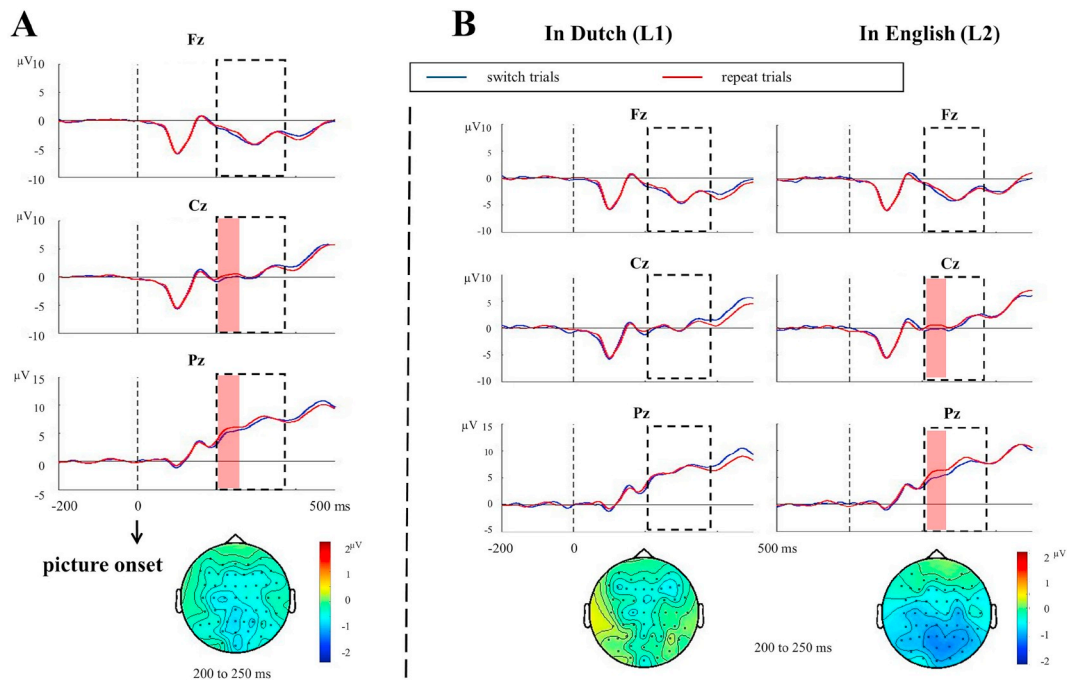


Fig. 5. (A) Stimulus-locked ERPs and topographies for switch vs. repeat trials (B) The same contrast (i.e., switch vs. repeat), when naming in L1 (Dutch) and in L2 (English) The time window used for testing the N2 effect (200–350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows and electrodes associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. We used the same time window in A and B (the window that was associated with the N2 effect in A) for depicting the topography for the sake of better comparability.

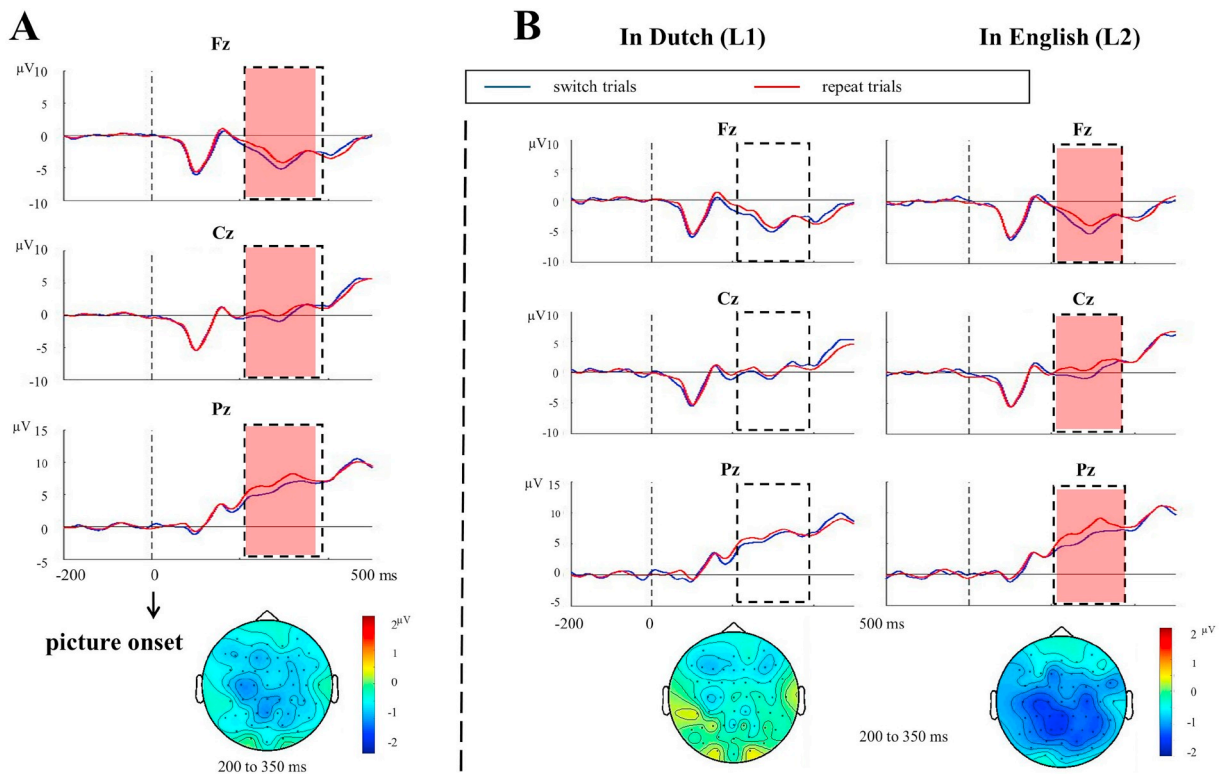


Fig. 6. (A) Stimulus-locked ERPs and topographies for switch cost (i.e., switch vs. repeat trials) after a short run (B) The same contrast (i.e., switch vs. repeat) when naming in L1 (Dutch) and in L2 (English). The time window used for testing the N2 effect (200–350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

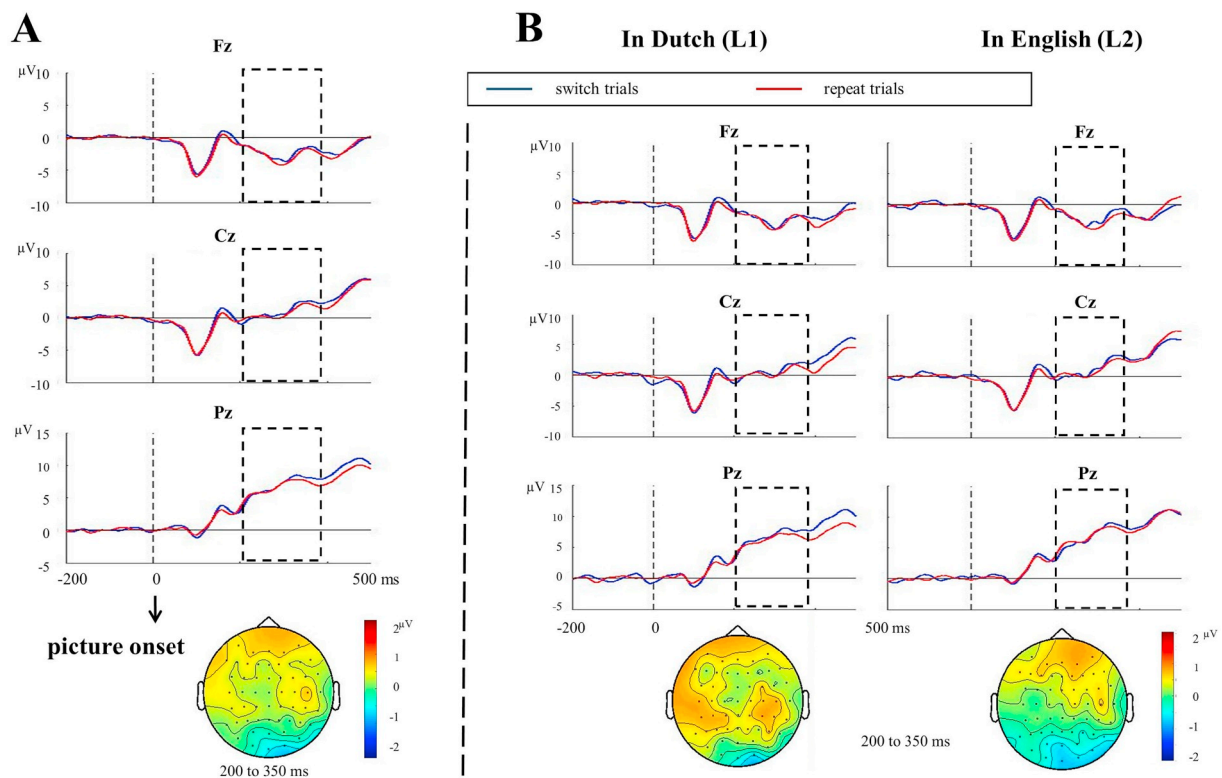


Fig. 7. (A) Stimulus-locked ERPs and topographies for switch cost (i.e., switch vs. repeat trials) after a long run (B) The same contrast (i.e., switch vs. repeat) when naming in L1 (Dutch) and in L2 (English). The time window used for testing the N2 effect (200–350 ms) is marked by an empty frame. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

and between run lengths. Fig. 5 shows the averaged ERPs and topographies for switch vs. repeat trials, and in each language.

The cluster-based permutation test revealed a larger N2 after switch compared to repeat trials ($p = .046$), with the effect being most pronounced at centro-posterior sites, from 200 to 250 ms post stimulus onset (Fig. 5A). When compared between languages, the switch cost N2 effect was present in L2 ($p = .022$), but not in L1 ($p = .655$). The effect in L2 was observed from 200 to 250 ms post stimulus onset, in centro-posterior sites (Fig. 5B). The difference between languages, however, did not reach significance ($p = .092$).

Switch costs following a short run. Fig. 6 shows the averaged ERPs and topographies for switch vs. repeat trials after a short run, and in each language.

The cluster-based permutation tests revealed a larger N2 following the switch after a short run compared to a repeat trial of early ordinal position ($p = .002$), with the effect being most pronounced from 200 to 340 ms post stimulus onset and widely spread over the scalp (Fig. 6A). When compared between languages, the switch cost N2 effect following a short run was present in the L2 ($p = .002$), but not in the L1 ($p = .158$). The effect in L2 was observed from 200 to 350 ms post stimulus onset, widely spread and more towards centro-posterior sites (Fig. 6B). The difference between languages, however, was not significant ($p = .282$).

Switch cost following a long run. Fig. 7 shows the averaged ERPs and topographies for switch vs. repeat trials after a long run, and in each language.

The cluster-based permutation test showed no N2 difference between switch trials after long runs and repeat trials of late ordinal position ($p = .320$; Fig. 7A). The switch cost after a long run, as reflected in the N2, was significantly different from that after a short run ($p = .006$), where a significant N2 effect was observed between trial types. When compared between languages, the switch cost N2 following a long run was neither present in the L2 ($p = .559$), nor in the L1 ($p = .432$). The difference between languages was also not significant ($p = .482$).

3.2.4. Summary

On switch trials, we observed a larger N2 after a short compared to a long run, with a wide-spread but more fronto-central scalp distribution. The N2 effect was only present in the L2, not in the L1, although the difference was not significant between languages. In contrast, no N2 difference was observed in early vs. late ordinal positions on repeat trials. Compared to the repeat trials, the N2 was enlarged on the switch trials. The switch cost was only present in the L2, not in the L1, with the N2 effect more pronounced at centro-posterior sites. The between-language difference was not significant, though. Moreover, the switch cost N2 was present only after a short run, but not after a long run.

4. Discussion

The current study investigated the dynamics of inhibitory control in bilingual speech production. We compared bilingual speakers' naming RTs and ERPs between short vs. long sequences of same-language trials (i.e., run length) in a language switching paradigm. Below, we first discuss the run length, switch cost, and reverse dominance effects in the RTs and ERPs. Next, we address the issue of whether the N2 effects reflect the application or overcoming of inhibition. Finally, the nature of top-down control, i.e., inhibition or enhancement, is discussed.

4.1. RT results

On the switch trials, speakers were faster after a long same-language run compared to a short run. This *run-length effect* was only present in the L1, not in the L2. This replicates the results reported by Zheng et al. (2018b), where a slightly different design was used. For example, in Zheng et al. a combination of cognate and noncognate items was used instead of only noncognate items, as in the current study. Moreover, the design and analyses in the previous study focused on error rates rather than RTs. It is important to note that a similar run-length effect has been

reported by Kleinman and Gollan (2018), which was, however, symmetrically present in both languages. In contrast, Meuter and Allport (1999) reported no changes in RTs on a switch as a function of the number of immediately preceding repeat trials. These inconsistent results might be due to differences in stimulus sets used in the studies: While the current study employed 40 unique pictures, participants repeatedly named nine pictures in the study by Kleinman and Gollan, and only single digits ranging from 1 to 9 in the study by Meuter and Allport.

To better compare the current study with the previous literature, we also calculated the switch costs, i.e., the difference between switch and repeat trials. As expected, the switch cost was larger after a short same-language run compared to a long run, indicating more effort to overcome the residual control after a short compared to a long run. On repeat trials, bilingual speakers were faster on trials in later than early ordinal position within a same-language run, suggesting increased bottom-up activation/priming of the target language. This also excludes the possibility that the run-length effect is due to expectation: If participants increasingly expect an upcoming switch (to the nontarget language) the longer a run gets, they should get increasingly slower on repeat trials in the current target language, rather than increasingly faster, within a same-language run. In other words, the expectation effect should facilitate the switch to the nontarget language, but impair the performance on the target language repetition, which is not the case in our data.

Moreover, during the task, speakers responded more quickly when naming in their L2 rather than L1, regardless of trial type. This so-called *reversed dominance effect* has often been observed before in mixed-language contexts (Christoffels et al., 2007; Costa and Santesteban, 2004; Verhoef et al., 2010), and is probably due to the global inhibition of L1 throughout the experiment in order to facilitate L2 production, while inhibition of the generally weaker L2 is less necessary. A similar phenomenon presumably also due to greater L1 relative to L2 inhibition is called *asymmetric switch cost*, namely, switching from the L2 to the L1 takes more time than the opposite switching direction (e.g., Meuter and Allport, 1999). Surprisingly, we found a slightly larger switch cost for switches to the L2 rather than to the L1. This might be due to the presence of the reversed dominance effect. Because of the frequent language switching during the task, the difficulty of speaking L1 on a switch trial (because of the residual control) seems to be carried over to the repeat trials. In other words, it appears that under the present conditions, switch cost asymmetry evolved into a reversed dominance effect. Under a reversed language dominance, L1 turns into the “weaker” language compared to L2, and therefore, switch cost asymmetries that hinge on this dominance difference become distorted. This explanation is supported by other studies showing a reversed dominance effect in absence of asymmetric switch costs (Christoffels et al., 2007; Verhoef et al., 2010, with numerically larger switch cost in L2 than in L1), or even with a reversed switch cost asymmetry as in our results (Declerck et al., 2015a).

Based on the behavioral results, we conclude that a target language gets more and more activated/primed with repetition and the demand of top-down control (inhibition of the nontarget language and/or enhancement of the target language) decreases over time. However, it remains unclear whether it is inhibition or enhancement that drives the run-length effect at switches.

4.2. ERP results

To examine the role of inhibitory control, we investigated the N2 component in ERPs. The N2 effect in language production is usually interpreted as reflecting application of inhibition (Jackson et al., 2001) or overcoming inhibition (Sikora et al., 2016). The former usually has a frontal or central scalp distribution (anterior N2), whereas the latter has a parietal or more posterior scalp distribution (posterior N2). On repeat trials, we observed no difference in the N2 amplitude between trials in early vs. late ordinal positions. Therefore, we failed to find evidence for

either accumulating or decreasing inhibition over a relative short number of language repetitions.

In contrast, on switch trials, a larger N2 was observed when switching after a short run compared to a long run, in line with our behavioral results. This N2 run-length effect had a broad scalp distribution with a more anterior rather than posterior topography. To better being able to compare our study with previous research (Jackson et al., 2001; Verhoef et al., 2010), we contrasted the ERPs between switch and repeat trials and compared the N2 switch cost between run lengths.

A small difference in N2 amplitude was observed between trial types, with more negative ERPs for switch trials compared to repeat trials, replicating earlier studies in language switching (e.g., Jackson et al., 2001; Verhoef et al., 2010; but see Christoffels et al., 2007). In line with the N2 analysis on switch trials as well as the RT results (i.e., larger switch costs after short compared to long runs), the switch cost was only present following a short run, not a long run. Different from the run-length effect in the N2 that we observed on switch trials, the topography of the “switch cost” N2 had a more posterior than frontal scalp distribution, resembling the one described in Verhoef et al. (2010) rather than Jackson et al. (2001). These results challenge the idea that inhibitory control is reactive (Green, 1998), which assumes that the amount of inhibition depends on the activation of a nontarget language. According to this assumption, a larger N2 should be expected at the switch following a long run, where the previous target language is more activated.

To obtain a more complete picture, we also compared the N2 as a function of trial type between languages. The switch cost N2 was only present in the L2, not in the L1, although the difference between languages was not significant. The results are in line with previous studies (Jackson et al., 2001; Verhoef et al., 2010), but the interpretation of this asymmetry remains unclear. It can reflect more inhibition of the L1 on switches to the L2 (Jackson et al., 2001) or disengaging from the stronger L1 on switches to the L2 compared to the opposite switching direction (Verhoef et al., 2010).

4.3. Inhibition vs. overcoming inhibition

Although the N2 effects observed in the current study usually have a widespread scalp distribution, sometimes the effect seemed to be more anterior (when contrasting switch trials following short vs. long runs), or more posterior (when contrasting the switch cost between run lengths and between languages). The anterior and posterior N2 topographies were hypothesized to reflect different control processes (i.e., inhibition vs. overcoming inhibition, respectively). However, a clear theoretical cut-off between the two is difficult. Switching is a complex process, involving multiple cognitive functions, such as shifting from the previous task to the target task, and inhibiting the nontarget task (Miyake et al., 2000). Thus, the N2 switch effect could be a combination of inhibition (anterior N2) and overcoming inhibition (posterior N2).

It is worth noting that the two studies in which a posterior N2 effect in language production has been reported did not use the same experimental paradigm (Sikora et al., 2016; Verhoef et al., 2010). In Sikora et al. (2016), a short phrase (e.g., “the fork”) needed to be inhibited while producing the long phrase (e.g., “the green fork”). Therefore, when switching back to the short phrase, more inhibition needed to be overcome (larger N2 effect) as compared to switching to the long phrase. Using the same logic, when switching to the L1, more residual inhibition needs to be overcome because the L1 is more strongly inhibited during L2 production. As a consequence, a larger N2 should be expected in switches to L1 than vice versa, which is neither the case in the current findings nor in previous studies (e.g., Jackson et al., 2001; Verhoef et al., 2010).

Actually, it is unclear in all these accounts what needs to be inhibited, to be overcome, or to be disengaged from. Previous literature diverges on this issue. Meuter and Allport (1999) speak about “disengagement” from the preceding language set (e.g., inhibiting the L1

and enhancing the L2 while speaking in the weaker L2) instead of disengaging from the previous target language (as proposed in Verhoef et al., 2010), which is very similar to the “overcoming inhibition” account (Sikora et al., 2016). In that scenario, the disengagement effort was reflected in the asymmetric switch costs (i.e., actively disengaging from the task set of preceding L2 repeat trials is more difficult than from the task set of preceding L1 repeat trials). If such a “disengagement” effort is reflected by the N2 amplitude as well, then one should expect a larger N2 effect in switching to L1 than to L2. Again, this is opposite to what has been found in Jackson et al. (2001), Verhoef et al. (2010), and in the current study.

However, the purpose of the current study was not to solve the inconsistency in the functional interpretation of the N2 in the literature, but to make use of our current knowledge of the N2 effect to investigate the run-length effect and the dynamics of inhibitory control during bilingual production. Although it is not clear whether the N2 in language switching reflects the application of inhibition or overcoming inhibition, we believe that it may serve as a more general index for control due to the high cognitive demand at a switch. As pointed out in the Introduction, the N2 has also been associated with the monitoring of conflicts (e.g., Christoffels et al., 2007; Morales et al., 2013). However, this interpretation is unlikely according to the current data. The longer RTs when switching to L1 than to L2 (hypothetically) indicate larger conflict in the former switching direction. Therefore, a larger N2 effect, as an index of conflict, should be expected when switching from L2 to L1 than vice versa, which is not the case in the current results.⁵

In the RT analysis, the run-length effect was present in the L1 rather than L2. In contrast, the N2 run-length effect was present only in the L2, although the between-language interaction was not significant. We speculate that this may be due to a negative relationship between the amount of inhibition, as reflected by the N2 amplitude, and the RT difference: A larger N2 at the switch to the L2 after a short L1 run suggests more inhibition compared to a switch after a long run. The successful application of inhibition causes a smaller increase of the RT compared to long runs. The opposite holds for switching to L1 after short vs. long runs: less inhibition (i.e., smaller N2 effect) and therefore a larger RT difference between short and long runs (see also Shao et al., 2014, for similar findings in monolingual language production).

4.4. Top-down control in language repetition: inhibition or enhancement?

The question of inhibition vs. overcoming inhibition aside, it also remains unclear whether it is inhibition or enhancement that drives the run-length effect. Given that there was no ERP evidence for either decreasing or accumulating inhibition for the repeat trials, the “residual control” to be overcome on the switch trials is more likely to be the residual effect of enhancement rather than inhibition (Allport and Wylie, 1999; Philipp et al., 2007). It is also possible that the difference in inhibitory control between early vs. late positions in a same-language run is not large enough to be visible in the ERPs on the repeat trials, but is large enough to make a difference at the switch where more top-down control is required, as reflected in the difference in effect between short vs. long runs. Inhibition of a nontarget language at a switch rather than during the repeated use of the target language is in line with the explanation of inhibition observed with n-2 language repetition costs (e.g., Declerck et al., 2015b; Guo et al., 2013; Philipp et al., 2007). N-2 language repetition costs refer to worse performance when switching back to a recently (two trials ago) abandoned language in trilingual switching (i.e., worse performance in ABA language

sequences than CBA sequences, where A, B, and C refer to different languages). When switching from language A to B (in an ABA sequence) or from C to B (in CBA), the previously used language A or C is strongly inhibited. Therefore, when switching back to A, it is harder to overcome the previous inhibition in the ABA sequence than the CBA sequence. In contrast, if the level of inhibition of A and C purely depends on the use of B, then no difference should be found in switch costs between ABA and CBA sequences.

In the study by Kleinman and Gollan (2018), naming RTs of the target picture increased within a block as a function of the number of unrelated pictures that had been named in the alternative language. This evidence for accumulating inhibition over time seems to contradict the run-length effect in the current study and in Zheng et al. (2018b) at first glance. However, such contradiction may not be too surprising and can even possibly be reconciled by taking into account different type of inhibitory processes, namely, more global, sustained inhibition versus more transient, local inhibition (Braver et al., 2003; De Groot and Christoffels, 2006). Whereas global inhibition refers to the suppression of one language as a whole, local inhibition works on a trial-by-trial basis. It is possible that within a same-language run, the inhibition of the nontarget language decreases due to the bottom-up priming of the target language. However, every time a switch to the alternative language occurs, the inhibition of the nontarget language needs to be brought back to a higher level. As a consequence, over the course of an entire language-mixing block, the overall inhibition of the nontarget language as a whole accumulates. Therefore, the effect of accumulative inhibition is not (merely) caused by the use of the target language, but rather by the frequent switching between languages. Future studies can investigate the effect of switching frequency on inhibition to further investigate the dynamics of language switching and repetition.

4.5. Summary

The current study explored the dynamics of inhibitory control during bilingual speech production by examining RTs and ERPs. We replicated the behavioral RT results as reported in Zheng et al. (2018b). The results suggest that top-down control (inhibition of the nontarget language and/or enhancement of the target language) is highest at a switch. With repeated use of the same language, the target language receives more and more bottom-up activation and RT decreases. As a consequence, top-down control gets reduced over time and becomes easier to be overcome, reflected in faster switching following a long than a short run. Correspondingly, we found a larger N2 effect following short same-language runs compared to long runs, indicating more control effort in the former case. In contrast, no difference in N2 was observed within a same-language run. Our ERP results suggest that bilingual speakers mainly inhibit a nontarget language at a switch rather than during the repeated use of a target language.

Data availability

Data are available from the Donders Repository at <http://hdl.handle.net/11633/aacbzlzc>.

CRediT authorship contribution statement

Xiaochen Zheng: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing - original draft, Writing - review & editing. **Ardi Roelofs:**

⁵ See the Discussion in Zheng et al. (2018a) for more elaborated arguments about the level of conflicts in the two switching directions.

Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Hasan Erkan:** Data curation, Formal analysis. **Kristin Lemhöfer:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgement

The authors would like to thank Jana Klaus, Manuel Augustin, Cesko Voeten, and Wouter Comes van Batenburg for their help during the data analysis and manuscript preparation.

Appendix A. Stimuli

	English name	Dutch name		English name	Dutch name
1	ant	mier	21	flower	bloem
2	arrow	pijl	22	frog**	kikker
3	ax	bijl	23	horse	paard
4	boot**	laars	24	key	sleutel
5	bottle	fles	25	knife	mes
6	bowl**	kom	26	leg**	been
7	box	doos	27	lettuce	sla
8	branch	tak	28	mirror	spiegel
9	brush	borstel	29	mountain**	berg
10	butterfly	vlinder	30*	mustache	snor
11	button	knoop	31	pig	varken
12	cage	kooi	32*	pillow	kussen
13	can	blik	33	plate	bord
14	candle	kaars	34	shower	douche
15	car**	auto	35	spoon	lepel
16	corn	mais	36	squirrel	eekhoorn
17	desk	bureau	37	tree**	boom
18	dog	hond	38	umbrella	paraplu
19	duck	eend	39	waiter**	ober
20	eye	oog	40	wall	muur

* These items were not in the original picture naming database, but added after the pilot study.

**These items occurred three instead of four times on a switch trial within a list.

Appendix B. Linear Mixed Effect Models

Overall analysis

```
glmer.RT.all
= glmer (RT ~ TrialType*Language*RL.OP + (1+ TrialType + Language + RL.OP|pNumber) + (1+ TrialType +
Language + RL.OP|PicNam), data = mydata.4RT.all, family = Gamma (link = "identity"), control = glmerControl
(optimizer = 'bobyqa'))
# full model fails to converge
```

Switch cost

```
# Switch cost in L1
glmer.RT.all.swista.L1
= glmer (RT ~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam), data = mydata.4RT.all [mydata.4RT.all
$Language == 'Dutch',], family = Gamma (link = "identity"), control = glmerControl (optimizer = 'bobyqa'))
# Switch cost in L2
glmer.RT.all.swista.L2
= glmer (RT ~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam), data = mydata.4RT.all [mydata.4RT.all
$Language == 'English',], family = Gamma (link = "identity"), control = glmerControl (optimizer = 'bobyqa'))

# Switch cost after a short run
glmer.RT.all.swista.short
= glmer (RT ~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam), data = mydata.4RT.all [mydata.4RT.all
$RL.OP == 'short',], family = Gamma (link = "identity"), control = glmerControl (optimizer = 'bobyqa'))
# Switch cost after a long run
glmer.RT.all.swista.long
= glmer (RT ~ TrialType + (1+ TrialType|pNumber) + (1|PicNam), data = mydata.4RT.all [mydata.4RT.all$RL.OP ==
'long',], family = Gamma (link = "identity"), control = glmerControl (optimizer = 'bobyqa'))
# full model fails to converge
```

Repeat trials

```
glmer.RT.repeat = glmer (RT ~ Language*OP + (1+Language + OP|pNumber) + (1+Language|PicNam), data =
mydata.4RT.repeat, family = Gamma (link = "identity"), control = glmerControl (optimizer = 'bobyqa'))
# full model fails to converge
```

Switch trials

```

glmer.RT.switch
= glmer(RT ~ Language*RL + (1+Language*RL|pNumber) + (1+Language + RL|PicNam), data = mydata.4RT.switch,
family = Gamma(link = "identity"), control = glmerControl(optimizer = 'bobyqa'))
# full model fails to converge
# Interaction
glm.RT.switch.L1
= glmer(RT ~ RL + (1 + RL|pNumber) + (1 + RL|PicNam), data = mydata.4RT.switch.[mydata.4RT.switch$Language
== "Dutch"], family = Gamma(link = "identity"), control = glmerControl(optimizer = 'bobyqa'))
glm.RT.switch.L2
= glmer(RT ~ RL + (1 + RL|pNumber) + (1 + RL|PicNam), data = mydata.4RT.switch [mydata.4RT.switch$Language
== "English"], family = Gamma(link = "identity"), control = glmerControl(optimizer = 'bobyqa'))

```

References

- Allport, A., Wylie, G., 1999. Task-switching: positive and negative priming of task set. In: Humphreys, G.W., Duncan, J., Treisman, A. (Eds.), *Attention, Space, and Action: Studies in Cognitive Neuroscience*. Oxford University Press, Oxford, pp. 273–296.
- Baayen, H., Piepenbrock, R., Gulikers, L., 1995. The CELEX Lexical Database [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia PA.
- Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, E., Dámico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., Tzeng, O., 2003. Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10, 344–380. <https://doi.org/10.3758/BF03196494>.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Boersma, P., Weenink, D., 2016. Praat: doing phonetics by computer [Computer program]. Retrieved 24 January 2016 from. <http://www.praat.org/>. Version 6.0.12.
- Braver, T.S., Reynolds, J.R., Donaldson, D.I., 2003. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* 39, 713–726. [https://doi.org/10.1016/S0896-6273\(03\)00466-5](https://doi.org/10.1016/S0896-6273(03)00466-5).
- Christoffels, I.K., Firk, C., Schiller, N.O., 2007. Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208. <https://doi.org/10.1016/j.brainres.2007.01.137>.
- Costa, A., Santesteban, M., 2004. Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *J. Mem. Lang.* 50, 491–511. <https://doi.org/10.1016/j.jml.2004.02.002>.
- De Groot, A.M.B., Christoffels, I.K., 2006. Language control in bilinguals: monolingual tasks and simultaneous interpreting. *Biling. Lang. Cognit.* 9, 189–201. <https://doi.org/10.1017/S1366728906002537>.
- Declerck, M., Philipp, A.M., 2015. A review of control processes and their locus in language switching. *Psychon. Bull. Rev.* 22, 1630–1645. <https://doi.org/10.3758/s13423-015-0836-1>.
- Declerck, M., Stephan, D.N., Koch, I., Philipp, A.M., 2015a. The other modality: auditory stimuli in language switching. *J. Cognit. Psychol.* 27, 685–691. <https://doi.org/10.1080/20445911.2015.1026265>.
- Declerck, M., Thoma, A.M., Koch, I., Philipp, A.M., 2015b. Highly proficient bilinguals implement inhibition: evidence from n-2 language repetition costs. *J. Exp. Psychol. Learn. Mem. Cognit.* 41, 1911. <https://doi.org/10.1037/xlm000138>.
- Falkenstein, M., Hoormann, J., Hohnsbein, J., 1999. ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychol.* 101, 267–291. [https://doi.org/10.1016/S0001-6918\(99\)00008-6](https://doi.org/10.1016/S0001-6918(99)00008-6).
- Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. <https://doi.org/10.1111/j.1469-8986.2007.00602.x>.
- Gollan, T.H., Ferreira, V.S., 2009. Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *J. Exp. Psychol. Learn. Mem. Cognit.* 35, 640–665. <https://doi.org/10.1037/a0014981>.
- Gollan, T.H., Goldrick, M., 2018. A switch is not a switch: syntactically-driven bilingual language control. *J. Exp. Psychol. Learn. Mem. Cognit.* 1, 1–14. <https://doi.org/10.1037/xlm0000462>.
- Gollan, T.H., Kleinman, D., Wierenga, C.E., 2014a. What's easier: doing what you want, or being told what to do? Cued versus voluntary language and task switching. *J. Exp. Psychol. Gen.* 143, 2167–2195. <https://doi.org/10.1037/a0038006>.
- Gollan, T.H., Schotter, E.R., Gomez, J., Murillo, M., Rayner, K., 2014b. Multiple levels of bilingual language control: evidence from language intrusions in reading aloud. *Psychol. Sci.* 25, 585–595. <https://doi.org/10.1177/0956797613512661>.
- Grainger, J., Dijkstra, T., 1992. On the representation and use of language information in bilinguals. In: Harris, R. (Ed.), *Cognitive Processing in Bilinguals*. North Holland, Amsterdam, pp. 207–220. [https://doi.org/10.1016/S0166-4115\(08\)61496-X](https://doi.org/10.1016/S0166-4115(08)61496-X).
- Green, D.W., 1998. Mental control of the bilingual lexico-semantic system. *Biling. Lang. Cognit.* 1, 67–81. <https://doi.org/10.1017/S1366728998000133>.
- Grosjean, F., 1998. Studying bilinguals: methodological and conceptual issues. *Biling. Lang. Cognit.* 1, 131–149. <https://doi.org/10.1017/S136672899800025X>.
- Grosjean, F., 1999. The bilingual's language modes. In: Nicol, J.L. (Ed.), *One Mind, Two Languages: Bilingual Language Processing*. Blackwell, Oxford, pp. 1–22.
- Guo, T., Liu, F., Chen, B., Li, S., 2013. Inhibition of non-target languages in multilingual word production: evidence from Uighur-Chinese-English trilinguals. *Acta Psychol.* 143, 277–283. <https://doi.org/10.1016/j.actpsy.2013.04.002>.
- Heikoop, K.W., Declerck, M., Los, S.A., Koch, I., 2016. Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Biling. Lang. Cognit.* 19, 921–927. <https://doi.org/10.1017/S1366728916000456>.
- Jackson, G.M., Swainson, R., Cunnington, R., Jackson, S.R., 2001. ERP correlates of executive control during repeated language switching. *Biling. Lang. Cognit.* 4, 169–178. <https://doi.org/10.1017/S1366728901000268>.
- Jodo, E., Kayama, Y., 1992. Relation of a negative ERP component to response inhibition in a Go/No-go task. *Clin. Neurophysiol.* 82, 477–482. [https://doi.org/10.1016/0013-4694\(92\)90054-L](https://doi.org/10.1016/0013-4694(92)90054-L).
- Kleinman, D., Gollan, T.H., 2018. Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition* 173, 115–132. <https://doi.org/10.1016/j.cognition.2018.01.009>.
- Lavric, A., Clapp, A., East, A., Elchlepp, H., Monsell, S., 2018. Is preparing for a language switch like preparing for a task switch? *J. Exp. Psychol. Learn. Mem. Cognit.* 45, 1224–1233. <https://doi.org/10.1037/xlm0000636>.
- Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>.
- Lo, S., Andrews, S., 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6, 1–16. <https://doi.org/10.3389/fpsyg.2015.01171>.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Mayr, U., Kliegl, R., 2003. Differential effects of cue changes and task changes on task-set selection costs. *J. Exp. Psychol. Learn. Mem. Cognit.* 29, 362–372. <https://doi.org/10.1037/0278-7393.29.3.362>.
- Meuter, R.F.I., Allport, A., 1999. Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. <https://doi.org/10.1006/jmla.1998.2602>.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognit. Psychol.* 41, 49–100. <https://doi.org/10.1006/cogp.1999.0734>.
- Morales, J., Gómez-Ariza, C.J., Bajo, M.T., 2013. Dual mechanisms of cognitive control in bilinguals and monolinguals. *J. Cognit. Psychol.* 25, 531–546. <https://doi.org/10.1080/20445911.2013.807812>.
- Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., Ridderinkhof, K.R., 2003. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognit. Affect Behav. Neurosci.* 3, 17–26. <https://doi.org/10.3758/CABN.3.1.17>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9. <https://doi.org/10.1155/2011/156869>.
- Philipp, A.M., Gade, M., Koch, I., 2007. Inhibitory processes in language switching: evidence from switching language-defined response sets. *Eur. J. Cognit. Psychol.* 19, 395–416. <https://doi.org/10.1080/09541440600758812>.
- R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schotter, E.R., Li, C., Gollan, T.H., 2019. What reading aloud reveals about speaking: regressive saccades implicate a failure to monitor, not inattention, in the prevalence of intrusion errors on function words. *Q. J. Exp. Psychol.* 174702181881948 <https://doi.org/10.1177/1747021818819480>.
- Severens, E., Van Lommel, S., Ratnckx, E., Hartsuiker, R.J., 2005. Timed picture naming norms for 590 pictures in Dutch. *Acta Psychol.* 119, 159–187.
- Shao, Z., Roelofs, A., Acheson, D.J., Meyer, A.S., 2014. Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Res.* 1586, 130–142. <https://doi.org/10.1016/j.brainres.2014.07.009>.
- Sikora, K., Roelofs, A., Hermans, D., 2016. Electrophysiology of executive control in spoken noun-phrase production: dynamics of updating, inhibiting, and shifting. *Neuropsychologia* 84, 44–53. <https://doi.org/10.1016/j.neuropsychologia.2016.01.037>.
- van Casteren, M., Davis, M.H., 2006. Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589. <https://doi.org/10.3758/BF03193889>.

- Verhoef, K.M.W., Roelofs, A., Chwilla, D.J., 2009. Role of inhibition in language switching: evidence from event-related brain potentials in overt picture naming. *Cognition* 110, 84–99. <https://doi.org/10.1016/j.cognition.2008.10.013>.
- Verhoef, K.M.W., Roelofs, A., Chwilla, D.J., 2010. Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *J. Cognit. Neurosci.* 22, 1832–1843. <https://doi.org/10.1162/jocn.2009.21291>.
- Zheng, X., Roelofs, A., Farquhar, J., Lemhöfer, K., 2018a. Monitoring of language selection errors in switching: not all about conflict. *PLoS One* 13, e0200397. <https://doi.org/10.1371/journal.pone.0200397>.
- Zheng, X., Roelofs, A., Lemhöfer, K., 2018b. Language selection errors in switching: language priming or cognitive control? *Language. Cognit. Neurosci.* 33, 139–147. <https://doi.org/10.1080/23273798.2017.1363401>.