



Bilingual switching between languages and listeners: Insights from immersive virtual reality

David Peeters^{a,b,c,*}

^a Tilburg University, Department of Communication and Cognition, TiCC, Tilburg, the Netherlands

^b Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

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



ARTICLE INFO

Keywords:

Bilingualism
Control
Language switching
Language production
Virtual reality

ABSTRACT

Perhaps the main advantage of being bilingual is the capacity to communicate with interlocutors that have different language backgrounds. In the life of a bilingual, switching interlocutors hence sometimes involves switching languages. We know that the capacity to switch from one language to another is supported by control mechanisms, such as task-set reconfiguration. This study investigates whether similar neurophysiological mechanisms support bilingual switching between different listeners, within and across languages. A group of 48 unbalanced Dutch-English bilinguals named pictures for two monolingual Dutch and two monolingual English life-size virtual listeners in an immersive virtual reality environment. In terms of reaction times, switching languages came at a cost over and above the significant cost of switching from one listener to another. Analysis of event-related potentials showed similar electrophysiological correlates for switching listeners and switching languages. However, it was found that having to switch listeners and languages at the same time delays the onset of lexical processes more than a switch between listeners within the same language. Findings are interpreted in light of the interplay between proactive (sustained inhibition) and reactive (task-set reconfiguration) control in bilingual speech production. It is argued that a possible bilingual advantage in executive control may not be due to the process of switching per se. This study paves the way for the study of bilingual language switching in ecologically valid, naturalistic, experimental settings.

1. Introduction

In everyday life, we often seamlessly switch between different interlocutors. A casual chat with one's office mate, for instance, may be followed by a conversation with one's boss, which in turn can be followed by a phone call with a friend. In the life of a bilingual, such switches between different listeners may or may not coincide with a switch in the language that is used. A large body of experimental work investigating the mechanisms that support bilinguals' capacity to exert control over their languages has indicated that switching *languages* comes at a cost (Bobb & Wodniecka, 2013; Declerck & Philipp, 2015; Kolers, 1966; Meuter & Allport, 1999). It is an open question, however, whether switching between different *listeners* in itself is costly as well. Furthermore, it is unclear whether similar neuropsychological mechanisms support the two types of switch (language switch vs listener switch). If we assume that largely similar, domain-general mechanisms are involved in supporting these different types of task switch, it is unknown whether the timing of their involvement differs. The present

study aimed to address these open issues by making use of immersive virtual reality technology to investigate behavioral and neural correlates of bilingual switching between life-size virtual listeners with similar or different language backgrounds. To set the stage for the present study, we will first discuss relevant theoretical and methodological considerations with regard to the experimental study of bilingual language switching.

1.1. Language switching: switch costs and reversed language dominance

Bilinguals' ability to switch languages on the basis of an external cue has been investigated extensively in studies using the *cued language-switching paradigm* (Costa & Santesteban, 2004; Declerck & Philipp, 2015; Macnamara, Krauthammer, & Bolgar, 1968; Meuter & Allport, 1999). In this paradigm, bilingual participants typically name digits or pictures in one language or the other as a function of an arbitrary cue, such as the background color of the computer monitor on which the critical stimuli appear. Trials on which the to-be-used language differs

* Corresponding author at: Tilburg University, Department of Communication and Cognition, P.O. Box 90153, NL-5000 LE, Tilburg, the Netherlands.

E-mail addresses: d.g.t.peeters@uvt.nl, david.peeters@mpi.nl.

from the previous trial (i.e., switch trials) typically result in longer response times than trials on which the to-be-used language is the same as on the previous trial (i.e., non-switch trials). It has been argued that this behavioral *switch cost* is the downstream consequence of competition between different task schemas (Green, 1998). The idea is that language-specific task schemas are in competition (Green & Abutalebi, 2013), such that when a participant is cued to name a picture in their second language (L2), the task schema related to naming a picture in their first language (L1) is inhibited, and *vice versa*. The inhibition is often assumed to also involve all lexical representations in the corresponding language (e.g., Peeters, Runnqvist, Bertrand, & Grainger, 2014). The time it takes to overcome this inhibition and reconfigure the task-set explains the switch cost (Green, 1998; Monsell, 2003). The language switch cost and its dominant theoretical interpretation nicely illustrate how control mechanisms must interact with linguistic memory representations in the brain of a bilingual.

Another theoretically interesting finding, revealed by the cued language-switching paradigm, is that unbalanced bilinguals are under certain circumstances capable of naming pictures faster in their weaker L2 than in their stronger, native L1 (e.g., Christoffels, Firk, & Schiller, 2007; Costa & Santesteban, 2004; Costa, Santesteban, & Ivanova, 2006; Gollan & Ferreira, 2009; Kleinman & Gollan, 2016; Liu, Timmer, Jiao, Yuan, & Wang, 2019; Peeters et al., 2014; Peeters & Dijkstra, 2018; Verhoef, Roelofs, & Chwilla, 2009; Verhoef, Roelofs, & Chwilla, 2010). This counter-intuitive *reversed language dominance* is commonly interpreted as evidence for sustained inhibition of the first language (e.g., Baus, Branzi, & Costa, 2015; Bobb & Wodniecka, 2013; Costa & Santesteban, 2004; Declerck & Philipp, 2015; Peeters & Dijkstra, 2018). It suggests that unbalanced bilinguals may proactively lower the availability of their L1 to a certain extent in mixed language contexts for a longer period of time, presumably to allow for less competition from their dominant L1, which leads to more efficient naming in their weaker L2 (Christoffels et al., 2007; Kroll, Bobb, Misra, & Guo, 2008; Peeters et al., 2014).

Results from the cued language-switching paradigm have thus been theoretically informative in suggesting that bilinguals may engage in reactive and transient task-set reconfigurations (as evident from the language switch cost) and proactive sustained inhibition of their L1 (as evident from reversed language dominance) in mixed language environments. These findings are further supported by the observation of mixing costs (Bobb & Wodniecka, 2013; Christoffels et al., 2007; Los, 1996), n-2 repetition costs (Declerck, Thoma, Koch, & Philipp, 2015; Philipp & Koch, 2009), order effects (Misra, Guo, Bobb, & Kroll, 2012; Peeters & Dijkstra, 2018), and preparation effects (Wu & Thierry, 2017). They are also in line with models of executive control that include both reactive and proactive control mechanisms (e.g., Braver, 2012). It is currently less clear, however, whether the proposed mechanisms play a role in the life of a bilingual outside the artificial, experimental lab setting.

1.2. Ecological validity of research findings on bilingual language production

Recent years have witnessed a plea for an increase in ecological validity of experimental paradigms in cognitive science (e.g., Hari, Henriksson, Malinen, & Parkkonen, 2015; Peeters, 2019; Schilbach et al., 2013), and the field of bilingualism is no exception (Blanco-Elorrieta & Pykkänen, 2018; Myers-Scotton, 2006; Peeters & Dijkstra, 2018). The common use of arbitrary cues, such as color, to induce language switches in the cued language-switching paradigm has led to a debate on whether artificially induced switches and their theoretical interpretation have any relevance for bilinguals switching languages in naturally occurring communication (Blanco-Elorrieta & Pykkänen, 2017; De Bruin, Samuel, & Duñabeitia, 2018; Kleinman & Gollan, 2016; Peeters et al., 2014). After all, in everyday life, language switches are typically not induced by the background color of computer monitors

(cf. Myers-Scotton, 2006).

Studies that consequently let bilingual participants *voluntarily* switch between their languages in the lab have in some cases observed switch costs comparable to traditional switch costs observed in cued language-switching paradigms, but in other cases found reduced switch costs or even no switch costs at all (Blanco-Elorrieta & Pykkänen, 2017; De Bruin et al., 2018; Gollan & Ferreira, 2009; Gollan, Kleinman, & Wierenga, 2014; Gross & Kaushanskaya, 2015; Jevtović, Duñabeitia, & de Bruin, 2019; Kleinman & Gollan, 2016; Reverberi et al., 2018). Findings on voluntary language switching may generalize mostly to dual language contexts in which language task schemas are not in competition (Green & Abutalebi, 2013; Heredia & Altarriba, 2001). In a bilingual language mode (Grosjean, 1998), bilinguals may indeed opportunistically use whatever lexical material comes to mind when communicating with other bilinguals that master the same languages (Green & Abutalebi, 2013). Nevertheless, in many language communities, bilinguals also often switch languages when consecutively encountering different monolingual interlocutors with varying language backgrounds.

Indeed, one salient external cue in everyday life that may induce a switch between languages is the language background of a speaker's intended addressee (Green & Abutalebi, 2013; Molnar, Ibáñez-Molina, & Carreiras, 2015). Results obtained with the cued language-switching paradigm should be particularly relevant in explaining the mechanisms involved in such naturally occurring situations. Research has shown that bilinguals may benefit from using the language associated with an interlocutor or their apparent cultural identity as a cue to use one language and not another (Blanco-Elorrieta & Pykkänen, 2015; Hartsuiker, 2015; Li, Yang, Scherf, & Li, 2013; Liu, Timmer, Jiao, Yuan, & Wang, 2019; Martin, Molnar, & Carreiras, 2016; Woumans et al., 2015). At the neurophysiological level, interlocutor identity has indeed been found to be represented in a sustained fashion during the planning stages that precede speech onset (Blanco-Elorrieta & Pykkänen, 2017).

A recent study asked bilingual participants to switch languages as a function of the language background of two monolingual virtual listeners. It was observed that switching languages as a function of monolingual listener identity comes at a cost that is similar to the switch cost found in studies that use arbitrary cues such as colors to induce a language switch (Peeters & Dijkstra, 2018). This finding was taken to support the ecological validity of the cued language-switching paradigm in indicating that traditional results may generalize to situations where bilinguals switch between different monolingual language modes (Grosjean, 1998). Nevertheless, as language switches in this previous study were always paired with a switch in the listener that was addressed, it raises the open question what the exact relation between *language switching* and *listener switching* is.

1.3. Switching languages versus switching listeners

The first aim of the present study is therefore to functionally disentangle language switches from listener switches. The majority of previous experimental studies using the cued language-switching paradigm used cue switches to induce language switches. The task switching literature, however, shows that a simple switch in cues alone may already lead to a cost in reaction time (Forstmann, Brass, & Koch, 2007; Mayr & Kliegl, 2003). In the domain of language switches, regardless of whether colors, flags, or listener identities are used as cues, language switches often involve cue switches, such that it has been difficult to separate the cost of switching languages from the cost of switching cues. Notable exceptions are recent studies in which a 2:1 cue-to-language mapping was used (e.g., De Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014, 2018; Heikooop, Declerck, Los, & Koch, 2016; Jevtović et al., 2019; Zheng, Roelofs, Erkan, & Lemhofer, 2019). These studies typically observe that switching languages comes at a cost over and above the cost associated with switching cues (see also Philipp & Koch, 2009). If these findings generalize to face-to-face situations in which

listeners are addressed on the basis of their monolingual language background, one would expect robust behavioral language switch costs over and above the cost of switching between different listeners.

The second aim of the present study is to investigate whether similar neurophysiological mechanisms support language switches and listener switches. This issue is relevant in the broader context of the widely debated extent of overlap between brain networks and neurocognitive mechanisms functionally involved in executive control and language control. Neuroimaging work indicates that the *language control network*, which subserves control operations when bilinguals produce (Abutalebi & Green, 2008, 2016; Luk, Green, Abutalebi, & Grady, 2012) and comprehend language (Peeters, Vanlangendonck, Rueschemeyer, & Dijkstra, 2019), overlaps with the network implicated in executive control more generally (e.g., Abutalebi et al., 2011; Blanco-Elorrieta & Pykkänen, 2016, 2017; Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2016; De Baene, Duyck, Brass, & Carreiras, 2015; De Bruin et al., 2014; Coderre, Smith, Van Heuven, & Horwitz, 2016; Garbin et al., 2010; Grant, Fang, & Li, 2015; Green & Abutalebi, 2013). The exact degree of overlap, however, is under debate and views differ in whether they posit only *partial* overlap (e.g., Calabria, Baus, & Costa, 2019; Magezi, Khateb, Mouthon, Spierer, & Annoni, 2012; Timmer, Calabria, Branzi, Baus, & Costa, 2018; Weissberger, Wierenga, Bondi, & Gollan, 2012) or *significant* overlap between *highly similar* brain circuits (e.g., Abutalebi & Green, 2016; De Baene et al., 2015; Plitsikas & Luk, 2016).

In the temporal domain, several event-related potential (ERP) components have been found to be sensitive to both language-related and domain-general control processes. A distinction can be made between relatively early components, such as the N2, that have been linked to general cognitive control processes such as task schema competition (Jackson, Swainson, Cunnington, & Jackson, 2001; Liu, Liang, Dunlap, Fan, & Chen, 2016) and later components, such as the late positive component (LPC), that in the case of bilingual language production research have often been linked to lexical processes (Jackson et al., 2001; Liu et al., 2016; Martin et al., 2013). The recent virtual reality study discussed above provided indirect ERP evidence for overlap between language control and executive control in bilinguals. A *switch negativity* in ERPs time-locked to picture onset was observed in unbalanced Dutch-English bilinguals naming pictures for two virtual listeners (Peeters & Dijkstra, 2018). This target-locked enhanced negativity for switch trials compared to non-switch trials, which was observed between 540–700 ms after picture onset (Peeters & Dijkstra, 2018), is reminiscent of the target-locked switch negativity found in the broader (non-linguistic) task switching literature (see Karayanidis & Jamadar, 2014, their Fig. 9.1b). The apparent similarity indirectly suggests that domain-general mechanisms support bilingual language switching (cf. Liu, Rossi, Zhou, & Chen, 2014).

Executive control and language control are broad notions that involve a variety of subcomponents, including goal maintenance, conflict monitoring, interference suppression, cue detection, response inhibition, and task (dis)engagement (Calabria, Costa, Green, & Abutalebi, 2018; Green & Abutalebi, 2013). It has traditionally been difficult to reliably compare non-linguistic executive control processes to linguistic control processes in the same bilingual population in tasks that are matched in all respects except for the involvement of language (Calabria et al., 2018; De Baene et al., 2015; Linck, Schwieter, & Sunderman, 2012; Soveri, Rodriguez-Fornells, & Laine, 2011; Timmer et al., 2018). The present study tackles this issue from a novel angle, by comparing instances of bilingual language production that differ in the number of task set reconfigurations necessary. When switching between two listeners that have the same monolingual language background, only *listener task schemas* (e.g., “Name picture for listener A” vs “Name picture for listener B”) need to be reconfigured. When switching between listeners that have a different monolingual language background, both *listener task schemas* and *language task schemas* will have to be reconfigured. A comparison between these two situations thus isolates the unique contribution of *language* switching from the more basic event of

switching from one listener to another. As both monolinguals and bilinguals alike commonly switch between different addressees in everyday life, language switches may rely more on bilingual *language* control processes than switches between listeners within the same language.

The third aim of the present study is to test whether the potential involvement of domain-general neuropsychological mechanisms supporting language switches and listener switches differs in the time-domain. Even if the same control mechanisms and the same underlying neural infrastructure would subservise different types of task switches (i.e., language switches vs listener switches), the timing of their involvement may differ, and abundant neuroimaging work relying on fMRI data in the study of bilingual language control has traditionally been more informative in the spatial domain than in the temporal domain. The onset of the late positive component, further discussed below, will be particularly informative in this temporal respect in the present study.

1.4. The present study

The bilingual language production experiment presented below makes use of immersive virtual reality technology to compare switches between listeners with a different language background (full switches) to switches between listeners with the same language background (listener switches) to trials on which the same listener is addressed in the same language as on the previous trial (non-switch trials). Unbalanced Dutch-English bilingual participants named pictures for four virtual addressees – two monolingual Dutch listeners and two monolingual English listeners – while for half of the participants their electroencephalogram (EEG) was continuously recorded. We know that this bilingual population shows symmetrical switch costs in a comparison of full switch trials to non-switch trials (Peeters & Dijkstra, 2018). If a substantial portion of these switch costs may be taken to reflect *language* control processes (Heikooop et al., 2016, p. 925), we should observe longer response times for full switch trials compared to listener switch trials. If *any* task schema reconfiguration is costly, also listener switches should come at a cost compared to non-switch trials. If, on the other hand, language switches were fully confounded by listener switches in previous work (e.g., Peeters & Dijkstra, 2018), no behavioral difference between full switches and listener switches should be observed.

At the neurophysiological level, the present study first aimed to replicate the enhanced switch negativity for full switch trials compared to non-switch trials in ERPs time-locked to picture onset (Peeters & Dijkstra, 2018). Critically, if language switches and listener switches share domain-general underlying control mechanisms, such as simple task schema reconfiguration, we should observe a modulation of the amplitude of this switch negativity for listener switches (compared to non-switch trials) as well. By zooming in on the onset and duration of the late positive component across L1 and L2 conditions, it will be possible to shed light on the timing and duration of the control processes involved during different types of switch. Additionally, assuming the replication of reversed language dominance (Peeters & Dijkstra, 2018) in the current study, enhanced positive ERP amplitude for the L2 compared to the L1 conditions should be observed (Liu et al., 2016; Peeters & Dijkstra, 2018), as it is sensitive to the selection of lexical items in an intended language (Martin et al., 2013). Timing differences in the onset of the late positive effect will therefore indicate that access to lexical representations may happen at different moments in time following language switches compared to following listener switches.

2. Method

2.1. Participants

Forty-eight native speakers of Dutch participated in the study.

Table 1

Participant characteristics. Participants performed the LexTALE lexical decision task in their L2 English, rated their proficiency in their L1 Dutch and L2 English, in addition to the frequency with which they read in both their languages. Ratings were made on a 7-point Likert scale. Values within parentheses are standard deviations. Behavioral data were collapsed across the two experiments in the behavioral analyses.

Experiment	Language	N	Age	LexTALE	Reading	Speaking	Reading Frequency	General Comprehension
Behavioral-only	L1	24	22.3 (2.6)	79.1 (10.8)	6.8 (0.7)	6.9 (0.3)	6.6 (0.6)	7.0 (0.2)
	L2				5.5 (1.1)	5.1 (1.1)	5.4 (1.8)	5.7 (1.0)
Behavioral + EEG	L1	24	23.0 (2.9)	79.3 (12.5)	6.8 (0.4)	6.8 (0.4)	6.3 (1.0)	7.0 (0.2)
	L2				5.5 (1.0)	5.3 (1.3)	4.7 (1.8)	5.8 (1.0)

Twenty-four (mean age 22.3, range 19–29 years old, 7 male, mean LexTALE score 79.1/100) took part in the behavioral-only experiment. Twenty-four others (mean age 23.0, range 18–29 years old, 6 male, mean LexTALE score 79.3/100) participated in the combined behavioral-EEG experiment. All participants were Dutch, studying in Nijmegen, right-handed (Oldfield, 1971), and Dutch was their single native language. They started learning their second language (L2) English between the ages of 8 and 12 in school. They had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided informed consent and were paid for participation. Data from four additional participants were not analyzed due to i) observation of a large number of picture naming errors (> 20 % of all trials) during the experiment ($n = 2$), ii) observation of a large number of EEG artifacts visible during the recording session ($n = 1$), and iii) accidentally having taken part in both experiments ($n = 1$). Sample size was *a priori* determined to be identical to a relevant previous study (Peeters & Dijkstra, 2018, their Experiments 3 and 4) that observed robust effects. Table 1 presents supporting information showing that participants qualified as unbalanced, but relatively proficient, Dutch-English bilinguals.

2.2. Stimuli and display

Forty single-object high quality color images were selected from a standardized online picture database (Moreno-Martínez & Montoro, 2012). The objects depicted in the pictures were taken from a number of different semantic categories, such as animals, body parts, clothing, and food (see Appendix A). Objects with English-Dutch cognate names (e.g., English *tomato*, Dutch *tomaat*) were avoided.

The experiment took place in a CAVE environment (Cruz-Neira, Sandin, & DeFanti, 1993), consisting of three large screens (255 cm × 330 cm, VISCON GmbH, Neukirchen-Vluyn, Germany) that were arranged in right angles. Two projectors (F50, Barco N.V., Kortrijk, Belgium) per screen projected a virtual environment onto each screen. Each set of two projectors showed two vertically displaced images which were overlapping in the middle of the screen. Thus, the complete display on each screen was only visible as combined overlay of the two projections. Infrared motion capture cameras (Bonita 10, Vicon Motion Systems Ltd, UK) and Tracker 3 software (Vicon Motion Systems Ltd, UK) were used for optical tracking. This exact virtual reality hardware setup has been described in detail before (Eichert, Peeters, & Hagoort, 2018; Heyselaar, Peeters, & Hagoort, 2019).

The experiment was programmed and run using Vizard software (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA). The virtual agents' speech was presented via two speakers (Logitech, US) that were located at the bottom edges of the screen facing the participants. The experiment took place in a virtual environment adapted from a stock environment produced by WorldViz ("room.wrl"). Wearing 3D shutter glasses, participants were immersed in this virtual room that contained a table on which a virtual screen resembling a computer monitor was presented. Pictures were presented one by one on this virtual monitor (Fig. 1). Standing behind the table, four virtual agents (all adapted from stock avatar "casual15_f_highpoly" produced by



Fig. 1. CAVE set-up used in the experiment. Participants named pictures in their L1 Dutch or their L2 English as a function of the language background of the virtual listener that looked at them. By wearing 3D shutter glasses, participants became immersed in the virtual environment.

WorldViz) served as relatively natural language cues in the experiment (see below).

2.3. Procedure

After providing informed consent, participants familiarized themselves with the pictures and their respective names in Dutch and English through a paper booklet. Before the start of the experiment, they were instructed to name pictures in either Dutch or English depending on the virtual agent that looked at them when a picture appeared on the virtual screen (see Fig. 1). They were told that two virtual agents ("Lotte" and "Marlou") only spoke and understood Dutch and that two other virtual agents ("Evelyn" and "Amie") only spoke and understood English. At the start of the experiment, the four virtual agents introduced themselves to the participant. Including lip sync and gaze at the participant, virtual agent Evelyn said "Hi, my name is Evelyn. I only speak English. Whenever I look at you, please tell me the name of the picture. Respond as quickly and accurately as possible". Virtual agent Amie then introduced herself in a similar way in English. Virtual agents Lotte and Marlou said the equivalent for Dutch in Dutch. The voices of the virtual agents were recorded from two native speakers of Dutch and two native speakers of English who matched the virtual agents in apparent age and ethnicity. Throughout the experiment, all virtual agents by default looked down. Critically, however, at the moment a picture appeared on the virtual screen, one of the virtual agents looked up and gazed at the participant, which served as a natural language cue. The physical appearance of the virtual agents in relation to their language identity (Dutch or English) and their position behind the virtual table were fully counterbalanced across participants. The main experiment was preceded by a practice set of six pictures that were not used in the main

experiment.

A trial started with a fixation cross presented on the virtual screen (2 s), followed by the presentation of a picture at the same moment one of the virtual agents looked at the participant. Speech onset was detected by a wireless Sennheiser microphone. Whenever the voice key was triggered, the picture disappeared from the screen and the screen remained blank for 3 s, after which the next trial started. The response deadline was set to 3 s. If the voice key was not triggered within 3 s from picture onset, the next trial started. Identical to an earlier study (Peeters & Dijkstra, 2018), an inter-trial interval of this length was opted for to allow for enough time to direct the virtual agents' gaze back to baseline and to allow for enough time for participants to blink between trials in the combined behavioral-EEG experiment. There were always at least ten trials between two presentations of the same picture. Pictures were presented in a pseudorandomized order, with a maximum of five subsequent trials from the same condition. There was a short break half way through the experiment.

Following the main experiment, participants performed the English LexTALE task (Lemhöfer & Broersma, 2012) on a laptop, filled out a language background questionnaire that was used to assess their self-reported proficiency in Dutch and English (Table 1), and were debriefed.

2.4. Design

NamingLanguage (L1 Dutch, L2 English) and TrialType (full switch, listener switch, non-switch) were within-subject factors in a 2×3 factorial design. A full switch trial was a trial on which both listener and language background of that listener differed compared to the previous trial (e.g., a switch from Dutch listener Lotte to English listener Evelyn). A listener switch trial was a trial on which there was a different listener who had to be addressed in the same language compared to the previous trial (e.g., a switch from Dutch listener Lotte to Dutch listener Marlou). A non-switch trial was a trial on which the same virtual agent was the listener compared to the previous trial. In order to keep the identity of the next listener unpredictable, each possible combination of two subsequently to-be-addressed virtual agents was included an equal number of 20 times in the experiment (cf. Heikoop et al., 2016). To keep the number of L1 trials identical to the number of L2 trials, this led to 160 full switch trials (80 per language), 80 listener switch trials (40 per language), and 80 non-switch trials (40 per language). Each unique picture was presented eight times (twice per listener; the same pictures in all conditions).

2.5. Behavioral analyses

Trials containing incorrect responses, no response within the response deadline, false starts, hesitations, and trials on which the microphone failed to record an RT were removed from the behavioral dataset (9.68 % of all trials). The first trial in the experiment and the first trial after the break were also excluded from further analysis. R (R Core Team, 2014), *lme4* (Bates, Maechler, Bolker, & Walker, 2015), and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) were used to test for possible switch costs and reversed language dominance in the picture naming reaction times (RTs) using linear mixed effects regression models (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). A logistic mixed effects regression analysis tested for possible switch costs and reversed language dominance in the picture naming error rates.

Three separate 2×2 contrasts were defined. First, we tested for switch costs and language dominance in the traditional way by including Language (L1, L2) and TrialType (full switch, non-switch) in the analysis. Second, we tested for the presence of within-language listener switch costs in a model comprising Language (L1, L2) and TrialType (listener switch, non-switch). Third, we tested whether any language switch costs were larger than listener switch costs in a model

comprising Language (L1, L2) and TrialType (listener switch, full switch). Behavioral data from the behavioral-only experiment and the combined behavioral-EEG experiment were collapsed, as there was no main effect of Experiment and no interaction effects between Experiment and the two independent variables in any of the RT analyses (see Table 3). Maximal random effects structures were used in the analysis of the RT data (cf. Barr et al., 2013; see Table 3). Note that Experiment and interaction terms including Experiment were not included as random slopes in the models as Experiment was a between-subject variable. Analyses of the error rate data were performed using random intercept models, as more complex models including random slopes did not converge.

2.6. Electrophysiological recording and analysis

Throughout the combined behavioral-EEG experiment (24 participants), the participant's EEG was recorded continuously from 59 active electrodes (Brain Products, Munich, Germany) held in place on the scalp by an elastic cap (Neuroscan, Singen, Germany). In addition to the 59 scalp sites, three external electrodes were attached to record participants electrooculogram (EOG), one below the left eye (to monitor for vertical eye movement/blinks), and two on the lateral canthi next to the left and right eye (to monitor for horizontal eye movements). Finally, one electrode was placed over the left mastoid bone and one over the right mastoid bone. The continuous EEG was recorded with a sampling rate of 500 Hz, a low cut-off filter of 0.01 Hz, and a high cut-off filter of 200 Hz. All electrode sites were referenced online to the electrode placed over the left mastoid and re-referenced offline to the average of the right and left mastoids.

Preprocessing and analyses of the electrophysiological data were carried out using Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). The raw EEG data was low-pass filtered offline at 40 Hz. Epochs from 100 ms preceding picture onset to 700 ms after picture onset were selected. The epochs were truncated at 700 ms, because after this point the earliest speech production artifacts started to contaminate the EEG signal. The 100 ms pre-stimulus period was used as a baseline. Trials defined as errors in the behavioral analyses and trials containing ocular or muscular artifacts were not taken into consideration in the averaging process (26.77 % of all trials – similar percentages across conditions). A total of 5394 trials entered the ERP analyses.

The event-related potential data were first analyzed using cluster-based permutation tests (Maris & Oostenveld, 2007) on the full epoch (0–700 ms). This non-parametric, data-driven approach has the advantage of controlling for the family-wise error rate that arises when an effect of interest is evaluated at multiple time points and electrodes (Maris & Oostenveld, 2007), thereby avoiding the multiple comparisons problem in the analysis of electrophysiological data (Maris, 2012). To describe the cluster-based permutation approach in short, for every data point (electrode by time), a simple dependent-samples *t* test comparing two conditions was performed. All adjacent data points (spatial or temporal) exceeding an alpha level of 0.05 were grouped into clusters. For all clusters (both positive and negative), the sum of the *t* statistics was used in the cluster-level test statistic. A null distribution was then calculated that assumed no difference between conditions (2000 randomizations, calculating the largest cluster-level statistic for each randomization), after which the actually observed cluster-level statistics were compared against this null distribution and clusters falling in the highest or lowest 2.5% percentile were considered significant (Bonferroni corrected; a *p* value < .025 reflects a significant effect). To test for onset differences in the effect of Language across types of trial (full switch, listener switch, non-switch), the temporal onset of significant clusters was taken as a suggestion for which time-windows to focus on in follow-up inferential testing (see below). Data from one participant was excluded from the EEG analysis due to a large number of EEG artifacts that had to be removed during the pre-processing stage.

Table 2

Mean reaction times (RT) in milliseconds and error rate proportions for the different conditions in the experiment. Only correct responses were included in the RT averages. Standard deviations are presented within parentheses.

Condition	Mean RT	Mean Error Rate
L1 Non-switch	1232 (328)	.058 (.23)
L1 Listener switch	1283 (358)	.083 (.28)
L1 Full switch	1325 (362)	.093 (.29)
L2 Non-switch	1130 (318)	.054 (.23)
L2 Listener switch	1160 (341)	.042 (.20)
L2 Full switch	1225 (352)	.071 (.26)

3. Results

3.1. Reaction time results

Table 2 and Fig. 2 provide the average picture naming RTs per condition in the experiment. In addition to showing reversed language dominance (i.e., faster RTs for L2 compared to L1), these data suggest that switching language comes at a cost over and above the cost of switching from one listener to another. Table 3 shows the results of three separate linear mixed effects analyses that were performed on these data to statistically test these result patterns.

First, tests for switch costs and language dominance were carried out in the traditional way by comparing full switch trials to non-switch trials. It was found that RTs were significantly longer for full switch trials ($M = 1275$ ms) compared to non-switch trials ($M = 1181$ ms). Moreover, RTs were significantly longer for L1 Dutch trials ($M = 1294$ ms) compared to L2 English trials ($M = 1194$ ms). No significant interaction between TrialType and Language was observed ($p = .80$). Hence, the predicted pattern of symmetrical switch costs and reversed language dominance was replicated in this bilingual population (Fig. 2, left panel).

Second, listener switches were compared to non-switch trials to test whether switching between different listeners within the same language comes at a cost. It was found that RTs were significantly longer for (within-language) listener switch trials ($M = 1220$ ms) compared to non-switch trials ($M = 1181$ ms). Moreover, RTs were again significantly longer for L1 Dutch trials ($M = 1257$ ms) compared to L2 English trials ($M = 1145$ ms). No significant interaction between TrialType and Language was observed ($p = .27$). Hence, it was established that switching listeners within the same language comes at a cost compared to continuing to speak in the same language to the same listener (Fig. 2, middle panel).

Third, most critically, within-language switches between different listeners (i.e., *listener switches*) were compared to across-language switches between different listeners (i.e., *full switches*). It was found that RTs were significantly longer for full switches ($M = 1275$ ms) compared to listener switches ($M = 1220$ ms). Moreover, RTs were again significantly longer for L1 Dutch trials ($M = 1311$ ms) compared to L2 English trials ($M = 1203$ ms). No significant interaction between TrialType and Language was observed ($p = .25$). Hence, these analyses confirm statistically that switching languages comes at a cost over and

above the significant cost of switching from one listener to another. In other words, language switch costs are partially, but not fully, driven by a switch in language cue (Fig. 2, right panel).

The present RT analyses did not exclude trials that were preceded by a trial on which an error was made. Excluding such trials from the analyses led to the same pattern of results as presented in this section. The outcome of these additional analyses is presented in the supplementary materials to this paper.

3.2. Error rate results

Errors were defined as trials containing incorrect responses, false starts, hesitations, and trials on which no response was given before the response deadline (3 s after picture onset). Table 2 provides the average error rate per condition in the experiment. Numerically, the error rates for the L1 conditions resembled the RT data, in that error rates (similar to RTs) increased with an increase in task set reconfigurations necessary (full switch > listener switch > non-switch). Also in L2, full switch trials yielded higher error rates compared to non-switch trials. L2 listener switches, however, elicited fewest errors. Similar to the RT analyses, three separate contrasts were defined to statistically test these result patterns.

First, switch costs and language dominance were tested for in the traditional way by comparing full switch trials to non-switch trials. A significant main effect of TrialType ($p = 1.21e^{-07}$) indicated that error rates for full switch trials ($M = .082$) were significantly higher than error rates for non-switch trials ($M = .056$). A significant main effect of Language ($p = .011$), indicated that error rates for L1 Dutch trials ($M = .081$) were significantly higher than error rates for L2 English trials ($M = .064$). There was no significant interaction between TrialType and Language ($p = .11$).

Second, listener switches were compared to non-switch trials. Overall, error rates for listener switch trials ($M = .063$) did not differ significantly ($p = .525$) from error rates for non-switch trials ($M = .056$). A significant main effect of Language ($p = 1.5e^{-05}$) indicated that error rates were higher for L1 Dutch trials ($M = .071$) than for L2 English trials ($M = .048$). In addition, a significant interaction between TrialType and Language ($p = .0004$) was observed. Follow-up analyses indicated that there was no effect of TrialType in the comparison of non-switch trials vs listener switch trials in L2 ($p = .06$). However, listener switch trials yielded a significantly higher error rate than non-switch trials in L1 ($p = .001$).

Third, within-language switches between different listeners (i.e., *listener switches*) were compared to across-language switches between different listeners (i.e., *full switches*). A significant main effect of TrialType was found ($p = 7.24e^{-06}$), indicating that, overall, full switches ($M = .082$) yielded more errors than listener switches ($M = .063$). A significant main effect of Language ($p = 2.85e^{-11}$) indicated that error rates were overall higher for L1 Dutch trials ($M = .090$) compared to L2 English trials ($M = .061$). There was, additionally, a significant interaction effect between TrialType and Language ($p = .009$). Follow-up analyses indicated that there was no effect of TrialType in the comparison of listener switches vs full switches in L1 ($p = .13$). However, full switches yielded a significantly

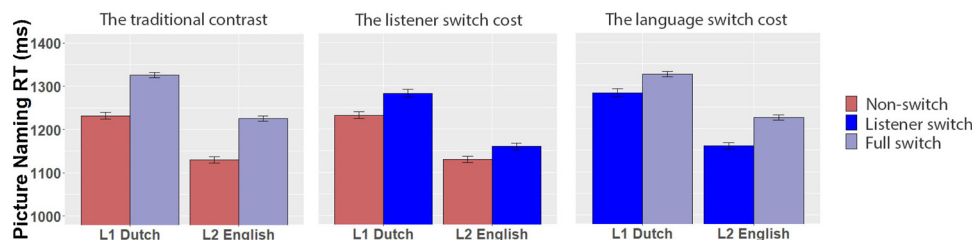


Fig. 2. Average picture naming response times provided for each of the three analyzed contrasts. Error bars represent standard errors around the mean. Note that each condition mean is represented in two different panels.

Table 3
Outcome of the linear mixed model analyses performed on the RT data.

1. The traditional comparison: full switch trials versus non-switch trials in L1 and L2

Model: ReactionTime ~ NamingLanguage*TrialType*Experiment + (1 + NamingLanguage*TrialType | Subject) + (1 + NamingLanguage*TrialType | Item)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1239.12	29.38	84.28	42.17	<2e-16***
NamingLanguage	-100.85	18.29	53.31	-5.51	1.05e-06***
TrialType	94.48	11.62	47.08	8.13	1.61e-10***
Experiment	-5.25	45.77	47.81	-0.12	0.91
NamingLanguage x TrialType	-5.64	21.64	39.38	-0.26	0.80
NamingLanguage x Experiment	21.55	19.19	47.81	1.12	0.27
TrialType x Experiment	23.55	14.96	48.42	1.58	0.12
NamingLanguage x TrialType x Experiment	6.55	24.20	236.34	0.27	0.79

2. Comparing non-switch trials to within-language listener-switches in L1 and L2

Model: ReactionTime ~ NamingLanguage*TrialType*Experiment + (1 + NamingLanguage*TrialType | Subject) + (1 + NamingLanguage*TrialType | Item)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1211.30	28.86	83.11	41.97	<2e-16***
NamingLanguage	-113.77	18.30	51.60	-6.22	9e-08***
TrialType	39.39	15.30	51.71	2.57	0.01*
Experiment	-10.98	45.74	47.83	-0.24	0.81
NamingLanguage x TrialType	-32.57	28.88	42.51	-1.13	0.27
NamingLanguage x Experiment	11.75	19.46	47.64	0.60	0.55
TrialType x Experiment	11.63	18.86	48.46	0.62	0.54
NamingLanguage x TrialType x Experiment	-14.35	28.84	97.35	-0.50	0.62

3. Comparing within-language listener switches to across-language listener switches in L1 and L2

Model: ReactionTime ~ NamingLanguage*TrialType*Experiment + (1 + NamingLanguage*TrialType | Subject) + (1 + NamingLanguage*TrialType | Item)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1258.70	30.58	82.25	41.17	<2e-16***
NamingLanguage	-116.72	18.40	50.27	-6.34	6.27e-08***
TrialType	55.35	15.05	44.62	3.68	<0.001***
Experiment	0.75	49.31	47.84	0.02	0.99
NamingLanguage x TrialType	26.32	22.38	40.30	1.18	0.25
NamingLanguage x Experiment	14.73	18.28	47.78	0.81	0.42
TrialType x Experiment	12.45	14.55	48.96	0.86	0.40
NamingLanguage x TrialType x Experiment	20.02	24.63	173.70	0.81	0.42

NamingLanguage (Dutch = -0.5; English = 0.5), TrialType (Non-switch = -0.5 and Full Switch = 0.5 in Model 1; Non-switch = -0.5 and Listener switch = 0.5 in Model 2; Listener switch = -0.5 and Full switch = 0.5 in Model 3), and Experiment (Behavioral experiment = -0.5; Behavioral + EEG experiment = 0.5) were deviation-coded. Significance codes: *** < .001; * < .05

higher error rate than listener switches in L2 ($p = 3.73e^{-06}$).

3.3. Electrophysiological results

Fig. 3 presents two ERP comparisons that tested for a switch negativity in the comparison of switch trials (full switches, listener switches) to non-switch trials. First, full switch trials were compared to non-switch trials, collapsed across Language (Dutch, English), to test whether this comparison replicated the switch negativity observed in earlier work (Peeters & Dijkstra, 2018). Cluster-based permutation tests on the full epoch (0–700 ms after stimulus onset) indeed revealed significantly ($p = .0005$) enhanced negative amplitude for full switch trials compared to non-switch trials, which was most pronounced between 452 and 700 ms after stimulus onset. Second, listener switch trials were compared to non-switch trials, collapsed across Language (Dutch, English), to test whether there was a similar switch negativity for listener switch trials compared to non-switch trials. Similar to the first comparison, cluster-based permutation tests on the full epoch (0–700 ms after stimulus onset) revealed significantly ($p = .0005$) enhanced negative amplitude for full switch trials compared to non-switch trials, which was most pronounced between 440 and 700 ms after stimulus onset. Finally, no statistically significant differences were observed in the comparison of full switches to listener switches (all p 's > .31).

Fig. 4 presents grand average waveforms for three more detailed comparisons in which L1 trials were compared to L2 trials for the three separate TrialType conditions (non-switch, listener switch, full switch). In all three comparisons, enhanced positive amplitude was observed for L2 trials compared to L1 trials in a time-window previously identified as showing a late positive component (e.g., Jackson et al., 2001; Liu et al., 2016; Martin et al., 2013). Cluster-based permutation testing suggested that the onset and duration of the effects of Language (enhanced positive amplitude for L2 vs L1) differed across the three comparisons. The effect of Language in the comparison of non-switch trials ($p = .0005$) was suggested to start 388 ms after stimulus onset. The effect of Language in the comparison of listener switch trials ($p = .0005$) was suggested to start 424 ms after stimulus onset. The effect of Language in the comparison of full switch trials ($p = .01$) was suggested to start 564 ms after stimulus onset. To statistically test whether the onset of these three effects of Language indeed differed significantly (cf. Sassenhagen & Draschkow, 2019), additional analyses were carried out. The onsets of the effects found for the three comparisons presented in Fig. 4, as provided by the cluster-based permutation tests, were taken as indications of which time-windows to focus on. Note that these additional analyses were relatively conservative, as averages were calculated based on data from all 59 recorded electrodes.

First, the average ERP value (in microvolt) over all 59 electrodes was calculated for the 388–424 ms time-window for each condition per

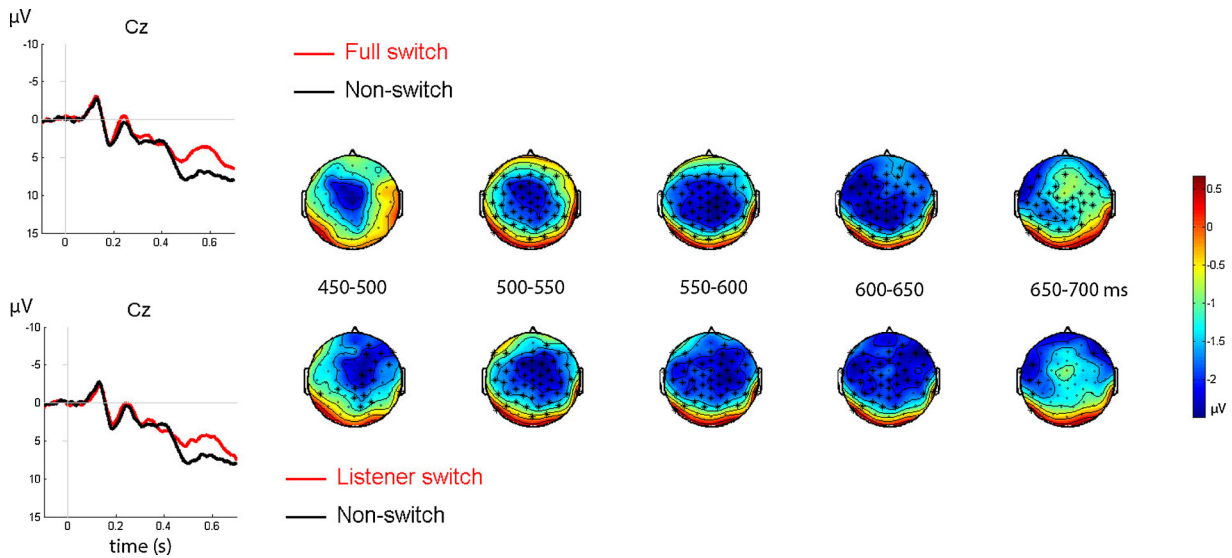


Fig. 3. Grand average waveforms time-locked to picture onset in a central electrode (Cz), comparing full switch trials to non-switch trials (top panel) and comparing listener switch trials to non-switch trials (bottom panel). Topographic plots show for each contrast the voltage difference between the two conditions over the scalp for 50 ms time-windows between 450 ms and 700 ms after stimulus onset.

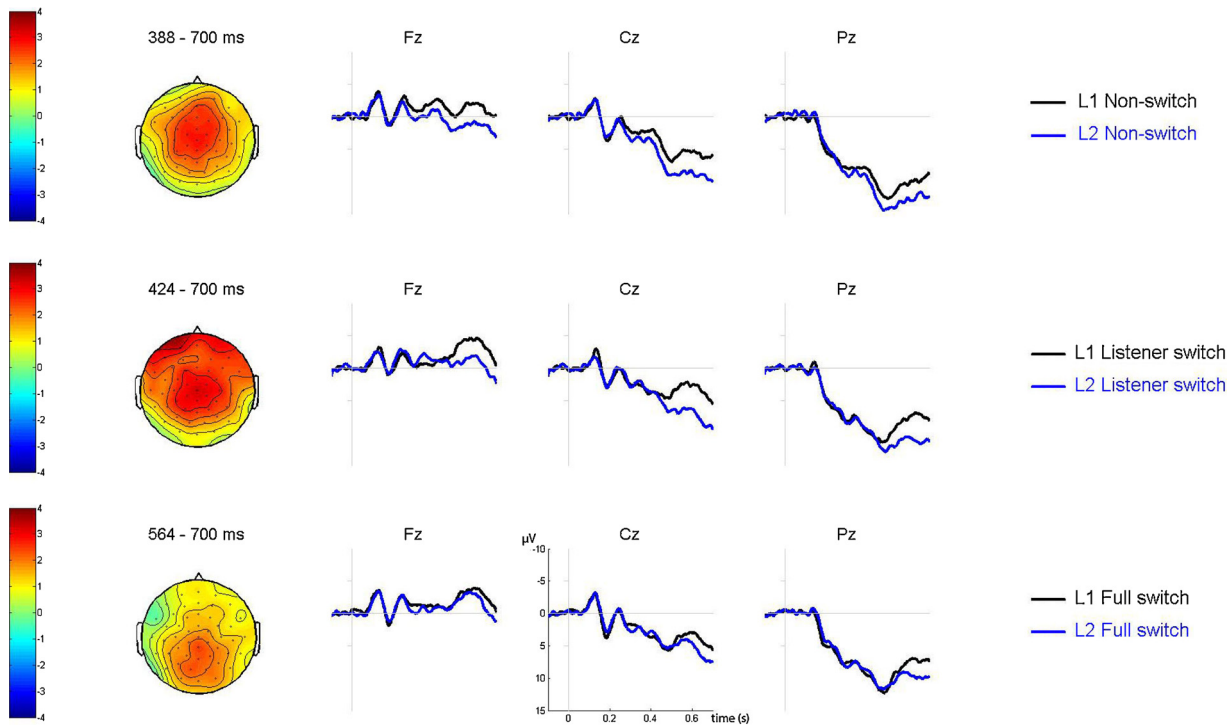


Fig. 4. Grand average waveforms time-locked to picture onset in three electrodes (Fz, Cz, Pz) for the three contrasts comparing L1 to L2 trials. The topographic plot shows for each contrast the voltage difference in microvolt between the two conditions over the scalp for the full time-window in which a significant difference was observed. Note that the onset of the effect of Language (L1 vs L2) differed as a function of the type of trial (non-switch, listener switch, full switch).

participant. Subtraction of the value for the L2 condition from the value for the L1 condition for each of the three contrasts presented in Fig. 4 yielded one difference score per contrast per participant. A simple paired *t* test confirmed that the average difference between the L1 and the L2 non-switch conditions (1.51 μV) was significantly higher ($p = .001$) than the average difference between the L1 and L2 full switch conditions (0.11 μV) in the 388–424 ms time-window. The difference between the difference score for the L1 and the L2 non-switch conditions and the difference score for the L1 and L2 listener switch conditions (0.75 μV) approached significance ($p = .079$) in this time-window. These results confirm that the ERP of effect of Language

started earlier in the non-switch comparison than in the full switch comparison.

Second, to test whether the onset of the effect of Language started significantly earlier in the comparison of the listener switch conditions compared to the comparison of the full switch conditions, similar analyses were carried out for the 424–564 ms time-window. To test for the emergence of an effect, this time-window was split into two equally long epochs. During the first half of this time-window (424–494 ms), no significant differences ($p = .222$) were observed between the average difference score for the two listener switch conditions (0.07 μV) and the average difference score for the two full switch conditions ($-0.14 \mu\text{V}$).

During the second half of this time-window (494–564 ms), the average difference score for the listener switch conditions (1.41 μV) differed significantly ($p = .001$) from the average difference score for the full switch conditions (0.36 μV).

Together, these findings suggest that lexical processes, which may differ between L1 trials and L2 trials as a consequence of the sustained inhibition of the L1, start at different points in time as a function of the number of task set reconfigurations involved.

4. Discussion

How do bilinguals manage to switch so efficiently between languages and how does a switch in language relate to a switch from one interlocutor to another? This study investigated behavioral and neural correlates of switching between different listeners, within and across languages. A group of 48 unbalanced Dutch-English bilinguals named pictures for two monolingual Dutch and two monolingual English virtual listeners in an immersive virtual reality environment. In terms of reaction times, switching languages came at a cost over and above the significant cost of switching from one listener to another. Furthermore, participants were faster and more accurate in their second language English compared to their native language Dutch. Analysis of event-related potentials showed similar electrophysiological correlates for switching listeners and switching languages. However, it was found that having to switch listeners and languages at the same time delays the onset of lexical processes more than a simple switch between two listeners with the same monolingual language background. We will now interpret these findings in light of the theoretical and methodological considerations put forward in the Introduction.

4.1. Language switching: switch costs and reversed language dominance

Studies using the cued language-switching paradigm have consistently observed that switching languages comes at a cost compared to using the same language on two consecutive trials. The large majority of previous studies using the paradigm, however, partially confounded language switches with cue switches, as language switches were typically induced by a switch in cues (but see De Bruin et al., 2014, 2018; Heikoop et al., 2016; Jevtović et al., 2019; Philipp & Koch, 2009; Zheng et al., 2019). Here participants named pictures for four life-size virtual listeners that used naturalistic gaze cues to induce language switches in a 2:1 cue-to-language mapping. It was observed that a significant portion of switch costs may be taken to reflect language control processes (cf. Heikoop et al., 2016). Switching listeners within the same language, moreover, came at a substantial cost as well.

These RT findings can be explained in a straightforward manner in terms of the number of task-set reconfigurations necessary to perform the naming task. The dominant interpretation of the language switch cost posits that language-specific task schemas are in competition (Green & Abutalebi, 2013). When a participant is cued to name a picture in their second language, the task schema related to naming a picture in their first language is inhibited, and *vice versa*. The time it takes to overcome the applied inhibition and reconfigure the task-set explains the switch cost (Green, 1998; Monsell, 2003). The results of the present study suggest that any task-set reconfiguration is costly. Inhibition of an active *listener task schema* (e.g., “Name picture for listener A”) when having to name a picture for listener B, within the same language, leads to a longer reaction time compared to when no switch is necessary. An additional, concurrent, switch in *language task schema* increases the RT even more. Switching languages and listeners at the same time is therefore more costly than switching listeners within the same language is.

Replicating earlier studies that tested different participant samples from the same population of unbalanced Dutch-English bilinguals (Peeters & Dijkstra, 2018), this study observed *symmetrical* switch costs and reversed language dominance in the RT data. In contrast, some

earlier studies testing unbalanced bilinguals observed larger switch costs for L1 than for L2 (e.g., Meuter & Allport, 1999). Such *asymmetrical* switch costs have been explained by assuming that the amount of inhibition applied to a language is proportional to its strength (Costa & Santesteban, 2004; Green, 1998; Meuter & Allport, 1999). As such, overcoming inhibition of one’s dominant language (often the L1) may take longer than overcoming inhibition of one’s weaker language (often the L2). The results of the present study confirm that unbalanced bilinguals do not necessarily show asymmetrical switch costs in the cued language-switching paradigm (Peeters & Dijkstra, 2018). Particularly when a relatively long response-cue interval is used (Ma, Li, & Guo, 2016), as in the present study, proactive control (as reflected by reversed language dominance) may play a more important role than reactive control (as typically reflected by asymmetrical switch costs). Indeed, by proactively lowering the availability of their L1, unbalanced bilinguals may experience less competition from their native language, which allows for more efficient naming in their weaker L2 (Christoffels et al., 2007; Kroll et al., 2008; Peeters & Dijkstra, 2018) and symmetrical switch costs across languages. The fact that we observed both slower RTs and higher error rates on L1 trials compared to L2 trials supports this explanation.

4.2. Switching languages versus switching listeners

This study aimed not only to functionally disentangle listener switches within and across languages, but also to investigate whether similar neurophysiological mechanisms support language switches and listener switches, and to test whether the potential involvement of domain-general neuropsychological mechanisms supporting language switches and listener switches differs in the time-domain. At a behavioral level, we have seen that switching languages comes at a cost over and above the significant cost of switching between different listeners. Nevertheless, similar task-set reconfiguration processes were proposed to be involved in both types of switch. Inspection of the switch negativity in the event-related potential data (Fig. 3) confirmed that language switches and listener switches were subserved by similar mechanisms. Both in terms of scalp distribution and latency, the switch negativity did not differ for listener switches within the same language compared to listener switches across languages. In line with the RT data and as suggested by visual inspection of ERPs and topoplots (Fig. 3), we predict that the switch negativity effect may actually last longer for listener switches across languages (Fig. 3, top panel) compared to listener switches within a language (Fig. 3, bottom panel). Nevertheless, based on these data, the underlying mechanism appears to be the same.

The observed neurophysiological similarities between listener switches within and across languages are theoretically relevant in the light of the broader debate on the degree of overlap between language control processes and executive control processes. A wide range of neuroimaging studies has found that the language control network overlaps with the network implicated in executive control more generally (e.g., Abutalebi et al., 2011; Blanco-Elorrieta & Pyllkänen, 2016, 2017; Branzi et al., 2016; De Baene et al., 2015; De Bruin et al., 2014; Coderre et al., 2016; Garbin et al., 2010; Grant et al., 2015; Green & Abutalebi, 2013). The present study confirms these findings from an ERP perspective by showing no electrophysiological differences in the switch negativity between a switch in listeners within *versus* across languages. Future studies using the present experimental paradigm could additionally include a condition in which a language switch is made in the absence of a listener switch, by including a bilingual listener that could be addressed on two subsequent trials in two different languages. This will allow for directly comparing the unique contribution of language switching to the unique contribution of listener switching.

The overlap between brain areas and mechanisms involved in *language control* and *executive control* more broadly has sometimes been used to argue that bilinguals, for instance when switching languages,

train their executive functions. As monolinguals by definition do not switch languages and do not need to inhibit one language when using the other, bilinguals may end up with better executive control skills than monolinguals. Such a bilingual advantage in executive functions has been widely debated (Abutalebi & Green, 2016; Bialystok, Kroll, Green, MacWhinney, & Craik, 2015; De Bruin, Treccani, & Della Sala, 2015; Dick et al., 2019; Paap & Greenberg, 2013) and the putative advantage has been most consistently observed in older adults, sometimes in children, and least in young adults (Antoniou, 2019, p. 398). The results reported in the present study are relevant in the context of this broader issue as it was found that switching listeners within a language elicits similar brain activity compared to switching listeners across languages. As monolinguals also commonly switch between different interlocutors, probably as often as bilinguals do, the present study suggests that it may not be the switching *per se* that leads to a potential bilingual advantage in executive functioning (cf. Verreyt, Woumans, Vandelandotte, Szmalec, & Duyck, 2016).

The observed reversed language dominance in the behavioral data suggests that our bilinguals proactively lowered the availability of their L1 in a sustained fashion. As a consequence, and in line with earlier work (Liu et al., 2016; Peeters & Dijkstra, 2018), the present study therefore consistently observed enhanced positive amplitude for L2 trials compared to L1 trials in the late positive component. It has been argued that the LPC is sensitive to the selection of a lexical item in an intended language, i.e. that it is related to the consequences of language control at the level of lexical representations (Martin et al., 2013, p. 7). In line with this proposal, the present LPC findings suggest that lexical processes (e.g., access to the mental lexicon, retrieval of lexical memory representations, and inhibition of representations in the non-target language) start at different moments in time as a function of the number of task-set reconfigurations involved on a particular trial. On non-switch trials, lexical processes may start as early as about 390 ms after stimulus onset. This onset is in line with the finding that it may take up to 300 ms for speakers to encode the interlocutor identity of their intended addressee (Blanco-Elorrieta & Pylkkänen, 2017). However, when having to make a switch in which listener to address while continuing to use the same language, the onset of the LPC effect was delayed. An additional, concurrent switch in which language to use delayed the onset of the LPC effect even more. In line with the reaction time results, these findings suggest that the number of required task-set reconfigurations may delay the onset of lexical processes, and as such eventually the articulation of the name of the pictured referent in the context-appropriate language. Hence the sustained inhibition of the L1 leads to greater difficulties in accessing L1 versus L2 representations, which is reflected in the amplitude differences in the LPC component across the two languages, which in turn is informative about the onset of lexical processes in the language production process.

It should be noted that the observed speech onset RT patterns across conditions strongly corresponded to the directionality and timing of amplitude differences in the ERP effects. In the individual comparisons presented in Figs. 3 and 4, conditions that yielded shorter RTs also elicited more positive amplitude in time-windows preceding speech onset. This raises the question whether the observed ERP differences can be ascribed to speech-related motor artifacts (Ganushchak, Christoffels, & Schiller, 2011; Porcaro, Medaglia, & Krott, 2015). It should be noted that the ERP analyses excluded (the minimal number of) trials on which speech was produced during the 700 ms epoch following picture onset. Furthermore, a recent study using a similar paradigm testing a different participant sample from the same bilingual population showed that faster RTs do not necessarily correlate with enhanced late positivity in ERPs preceding speech onset (Peeters & Dijkstra, 2018). In that study, trials in single language blocks consistently yielded faster RTs than non-switch trials in mixed language blocks. Nevertheless, in the ERP comparison of these conditions, enhanced positive ERP amplitude was found for the (slower) non-switch trials. In the light of this absence of a consistent correlation between RT

pattern and ERP effect directionality across similar studies, it is considered highly unlikely that the ERP effects observed in the present study can be due to motor activity.

Finally, it should be noted that the observed ERP effects of switching (Fig. 3) and the observed ERP effects of language (Fig. 4) overlap in time. This strongly suggests that, once listener identity and corresponding language background have been retrieved during the first ~390 ms of a trial in our paradigm, processes related to making the switch (e.g., task-set reconfigurations) and processes related to retrieval of linguistic memory representations in the bilingual mental lexicon occur, at least partially, in parallel. It has been hypothesized that, during language *comprehension*, temporo-parietal regions and frontal regions of the brain form a recurrent network through which activation reverberates (Hagoort, 2017). Similarly, during bilingual language *production*, temporal regions storing lexical information may be in continuous interaction with the language control network (Abutalebi & Green, 2016) to allow for the most efficient retrieval and production of context-appropriate spoken words (cf. Coderre et al., 2016).

4.3. Ecological validity of research findings on bilingual language production

The majority of previous studies using the cued language-switching paradigm have removed the act of bilingual language production from its natural, multimodal context of face-to-face communication. Artificially induced language switches made in front of a computer monitor indeed do not resemble the switches that bilinguals make in naturally occurring communication (Myers-Scott, 2006). The present study should be seen as part of a broader movement that aims to create ecologically valid experimental paradigms to study human cognition (e.g., Schilbach et al., 2013; Willems, 2015). Recent advances in the availability of immersive virtual reality technology now make it possible to design experimental paradigms in which high ecological validity and strict experimental control go hand in hand (Pan & Hamilton, 2018; Peeters, 2019; Tromp, Peeters, Meyer, & Hagoort, 2018). In the study of bilingual language production, this for instance now allows for perfectly matching different interlocutors within the same study in all aspects except their language background (Peeters & Dijkstra, 2018). Similar approaches using human confederates rather than virtual agents cannot reach such degree of experimental control (Kuhlen & Brennan, 2013). It is indeed hard to imagine how the use of four actual human listeners would have allowed for an equally high degree of millisecond-precise experimental control, as human listeners will not be able to produce identical non-verbal behavior across conditions and participants. Moreover, by using the same virtual agents in different studies and different research labs, the replicability of findings in the domain of face-to-face communication can now be assessed more reliably than before (Pan & Hamilton, 2018).

As a next step forward towards experimentally mimicking naturally occurring instances of bilingual language switching, future studies could replace the picture-naming component used in language production paradigms by a situation in which bilinguals refer to various 3D objects in realistic 3D environments for speakers with different language backgrounds. Such step-by-step improvements of existing paradigms will help moving towards neurocognitive theories of language that do justice to the multimodal richness and dynamics of naturally occurring human communication (Hari et al., 2015; Knoeferle, 2015). Moreover, the advantage of a step-by-step approach away from an existing paradigm is that it allows one to exactly pinpoint which factor drives a potential deviation in result patterns compared to results obtained in a traditional experimental paradigm. Future studies using virtual and augmented reality will show whether robust experimental findings (switch costs, reversed language dominance) fully generalize to different types of naturalistic situations.

5. Conclusion

To conclude, the present study shows the importance of the interplay between proactive (sustained inhibition) and reactive (task-set reconfiguration) control mechanisms in bilingual speech production. At a neurophysiological level, the mechanisms involved in switching between listeners did not differ within versus across languages. Furthermore, the observed results suggest that a possible bilingual advantage in executive control may not be due to the process of switching per se. The successful use of immersive virtual reality as an experimental method in bilingualism research paves the way for the experimental study of bilingual language switching in ecologically valid naturalistic settings.

Acknowledgments

The author would like to thank Angela de Bruin and Daniel Kleinman for valuable feedback on an earlier draft of this paper, Jeroen Derks for programming the experiment, Albert Russel, Johan Weustink, and Reiner Dirksmeyer for technical support, and Iris Schmits and Birgit Knudsen for help in testing participants. The author was partially supported by NWO Veni grant 275-89-037.

Appendix A. Picture names used in the experiment

L1 Dutch picture names. aardappel, aardbei, armband, aubergine, been, berg, bijl, boom, broek, busje, duif, eend, geweer, golven, haai, handschoen, jas, kerk, ketting, kip, meeuw, mier, neushoorn, oog, paard, perzik, pijl, potlood, schep, schildpad, slang, spijker, stoel, ui, vlieg, vliegtuig, vlinder, wolk, wortel, zaag.

L2 English picture names. ant, arrow, axe, bracelet, butterfly, carrot, chair, chicken, church, cloud, coat, duck, eggplant, eye, fly, glove, gun, horse, leg, mountain, nail, necklace, onion, peach, pencil, pigeon, plane, potato, rhino, saw, seagull, shark, shovel, snake, strawberry, tree, trousers, turtle, van, waves.

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.cognition.2019.104107>.

References

- Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, 23(4), 557–582.
- Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, 19(4), 689–698.
- Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, 22(9), 2076–2086.
- Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, 5, 395–415.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Baus, C., Branzi, F. M., & Costa, A. (2015). On the mechanism and scope of language control in bilingual speech production. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 508–526). Cambridge, UK: Cambridge University Press.
- Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*, 26(6), 944–946.
- Blanco-Elorrieta, E., & Pykkänen, L. (2015). Brain bases of language selection: MEG evidence from Arabic-English bilingual language production. *Frontiers in Human Neuroscience*, 9, 27.
- Blanco-Elorrieta, E., & Pykkänen, L. (2016). Bilingual language control in perception versus action: MEG reveals comprehension control mechanisms in anterior cingulate cortex and domain-general control of production in dorsolateral prefrontal cortex. *Journal of Neuroscience*, 36(2), 290–301.
- Blanco-Elorrieta, E., & Pykkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, 37(7), 9022–9036.
- Blanco-Elorrieta, E., & Pykkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences*, 22(12), 1117–1126.
- Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, 25(5), 568–585.
- Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex*, 26(6), 2367–2380.
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113.
- Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences*, 1426(1), 221–235.
- Calabria, M., Baus, C., & Costa, A. (2019). Cross-Talk between language and executive control. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 447–466). Hoboken, NJ: Wiley-Blackwell.
- Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, 1147, 192–208.
- Coderre, E. L., Smith, J. F., Van Heuven, W. J., & Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Bilingualism: Language and Cognition*, 19(3), 471–488.
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491–511.
- Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1057–1074.
- Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of The CAVE. *Proceedings of the 20th annual conference on computer graphics and interactive techniques*, 135–142.
- De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. *Journal of Cognitive Neuroscience*, 27(9), 1752–1765.
- De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: fMRI evidence from trilingual speakers. *NeuroImage*, 90, 348–359.
- De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 99–107.
- De Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, 103, 28–43.
- Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, 22(6), 1630–1645.
- Declerck, M., Thoma, A., Koch, I., & Philipp, A. M. (2015). Highly proficient bilinguals implement inhibition—Evidence from N-2 repetition costs when switching between three languages. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 41, 1911–1916.
- Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., et al. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behaviour*, 3, 692–701.
- Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods*, 50(3), 1102–1115.
- Forstmann, B. U., Brass, M., & Koch, I. (2007). Methodological and empirical issues when dissociating cue-related from task-related processes in the explicit task-cuing procedure. *Psychological Research*, 71(4), 393–400.
- Ganushchak, L., Christoffels, I., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology*, 2, 208.
- Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodríguez-Pujadas, A., Belloch, V., et al. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, 53(4), 1272–1278.
- 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 640–665.
- Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, 143(6), 2167–2195.
- Grant, A. M., Fang, S. Y., & Li, P. (2015). Second language lexical development and cognitive control: A longitudinal fMRI study. *Brain and Language*, 144, 35–47.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67–81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530.
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131–149.
- Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English-Spanish bilingual children. *Journal of Cognitive Psychology*, 27(8), 992–1013.
- Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience and Biobehavioral Reviews*, 81, 194–204.

- Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, 88(1), 181–193.
- Hartsuiker, R. J. (2015). Visual cues for language selection in bilinguals. In F. Huettig (Ed.), *Attention and vision in language processing* (pp. 129–145). India: Springer.
- Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, 19(5), 921–927.
- Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, 10(5), 164–168.
- Heyselaar, E., Peeters, D., & Hagoort, P. (under review). Do we predict upcoming speech content in naturalistic environments?
- Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, 4(2), 169–178.
- Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2019). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, 1–13.
- Karayanidis, F., & Jamadar, S. D. (2014). Event-related potentials reveal multiple components of proactive and reactive control in task switching. In J. A. Grange, & G. Houghton (Eds.), *Task switching and cognitive control* (pp. 200–236). New York, NY: Oxford University Press.
- Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, 27(5), 700–714.
- Knoeferle, P. (2015). Language comprehension in rich non-linguistic contexts: Combining eye tracking and event-related brain potentials. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 77–100). Cambridge: Cambridge University Press.
- Kolers, P. A. (1966). Reading and talking bilingually. *The American Journal of Psychology*, 79(3), 357–376.
- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, 128(3), 416–430.
- Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20(1), 54–72.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343.
- Li, Y., Yang, J., Scherf, K. S., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language*, 127(3), 452–462.
- Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, 15(3), 651–662.
- Liu, H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS One*, 9(10), e110887.
- Liu, C., Timmer, K., Jiao, L., Yuan, Y., & Wang, R. (2019). The influence of contextual faces on bilingual language control. *Quarterly Journal of Experimental Psychology*, 72(9), 2313–2327.
- Liu, H., Liang, L., Dunlap, S., Fan, N., & Chen, B. (2016). The effect of domain-general inhibition-related training on language switching: An ERP study. *Cognition*, 146, 264–276.
- Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, 94(2), 145–188.
- Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, 27(10), 1479–1488.
- Ma, F., Li, S., & Guo, T. (2016). Reactive and proactive control in bilingual word production: An investigation of influential factors. *Journal of Memory and Language*, 86, 35–59.
- Macnamara, J., Krauthammer, M., & Bolgar, M. (1968). Language switching in bilinguals as a function of stimulus and response uncertainty. *Journal of Experimental Psychology*, 78(2), 208–215.
- Magezi, D. A., Khateb, A., Mouthon, M., Spierer, L., & Annoni, J. M. (2012). Cognitive control of language production in bilinguals involves a partly independent process within the domain-general cognitive control network: Evidence from task-switching and electrical brain activity. *Brain and Language*, 122(1), 55–63.
- Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, 49(4), 549–565.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Martin, C. D., Strijkers, K., Santesteban, M., Escera, C., Hartsuiker, R. J., & Costa, A. (2013). The impact of early bilingualism on controlling a language learned late: An ERP study. *Frontiers in Psychology*, 4, 815.
- Martin, C. D., Molnar, M., & Carreiras, M. (2016). The proactive bilingual brain: Using interlocutor identity to generate predictions for language processing. *Scientific Reports*, 6, 26171.
- Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 362–372.
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40(1), 25–40.
- Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, 67(1), 224–237.
- Molnar, M., Ibáñez-Molina, A., & Carreiras, M. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language*, 81, 91–104.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140.
- Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS One*, 7(5), e37527.
- Myers-Scotton, C. (2006). Natural codeswitching knocks on the laboratory door. *Bilingualism: Language and Cognition*, 9(2), 203–212.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Oostenveld, R., Fries, P., Maris, E., & Schoffeleers, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 1–9.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258.
- Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395–417.
- Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review*, 26(3), 894–900.
- Peeters, D., & Dijkstra, T. (2018). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition*, 21(5), 1035–1061.
- Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 284–292.
- Peeters, D., Vanlangendonck, F., Rueschemeyer, S. A., & Dijkstra, T. (2019). Activation of the language control network in bilingual visual word recognition. *Cortex*, 111, 63–73.
- Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching among languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 35, 1187–1195.
- Pliatsikas, C., & Luk, G. (2016). Executive control in bilinguals: A concise review on fMRI studies. *Bilingualism: Language and Cognition*, 19(4), 699–705.
- Porcaro, C., Medaglia, M. T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. *NeuroImage*, 105, 171–180.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reverberi, C., Kuhlen, A. K., Seyed-Allaei, S., Greulich, R. S., Costa, A., Abutalebi, J., et al. (2018). The neural basis of free language choice in bilingual speakers: Disentangling language choice and language execution. *NeuroImage*, 177, 108–116.
- Sassenshagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), e13335.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414.
- Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology*, 2, 183.
- Timmer, K., Calabria, M., Branzi, F. M., Baus, C., & Costa, A. (2018). On the reliability of switching costs across time and domains. *Frontiers in Psychology*, 9, 1032.
- Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, 50(2), 862–869.
- Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, 110(1), 84–99.
- Verhoef, K. M., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, 22(8), 1832–1843.
- Verreyt, N., Woumans, E. V. Y., Vandelanotte, D., Szmalc, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, 19(1), 181–190.
- Weissberger, G. H., Wierenga, C. E., Bondi, M. W., & Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging*, 27(4), 959–974.
- Willems, R. M. (2015). *Cognitive neuroscience of natural language use*. Cambridge: Cambridge University Press.
- Woumans, E., Martin, C. D., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R. J., et al. (2015). Can faces prime a language? *Psychological Science*, 26(9), 1343–1352.
- Wu, Y. J., & Thierry, G. (2017). Brain potentials predict language selection before speech onset in bilinguals. *Brain and Language*, 171, 23–30.
- Zheng, X., Roelofs, A., Erkan, H., & Lemhofer, K. (2019). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study. *bioRxiv*, 596445.