Accepted refereed manuscript of: Kuipers JR & Westphal K (2021) Auditory processing and high task demands facilitate the bilingual executive control advantage in young adults. *Journal of Neurolinguistics*, 57, Art. No.: 100954. https://doi.org/10.1016/j.jneuroling.2020.100954

Auditory processing and high task demands facilitate the bilingual executive control advantage in young adults

Jan R. Kuipers[1*]

Karla H. Westphal[1]

1. University of Stirling

School of Natural sciences

Psychology

*Corresponding author

Address for correspondence:

Dr. Jan Kuipers

University of Stirling

School of Natural sciences

Psychology

FK9 4LA Stirling,

United Kingdom

E-mail: jk28@stir.ac.uk

Abstract

Although bilingual children and elderly have been observed to outperform monolinguals in typical executive control tasks, this bilingual advantage is not consistently found in the young adult population. Proponents of the bilingual executive control advantage argue the reason for this is that task demands in the typical tasks used are not high enough, since young adults perform at ceiling level, whereas critics of the effect argue it has benefited from publication bias. Here we test the task-load hypothesis using a standard and a difficult version of the arrow-flanker task and identify stimulus processing characteristics underlying greater bilingual executive control. We increased task demands by using an "Opposite" task in which participants were to respond to the central arrow indicating its opposite direction whilst a task cue indicated which task was to be performed at each trial. Further increase in task difficulty was expected to arise from reducing the task preparation time by using different stimulus-onset-asynchronies between cue and target stimuli. As predicted, we observed no language group differences in the normal flanker task, whereas bilinguals displayed less errors than monolinguals and were less hampered by the difficult task than monolinguals when auditory task cues were used. Event-related potentials (ERPs) revealed that the bilinguals' conflict monitoring response occurred much earlier than the monolinguals' when the task cue was auditory but less so when the cue was visual. Indeed, bilinguals appeared to prioritize the cue signal when it was auditory, but not when it was visual. Further ERP results showed bilinguals displayed greater attentional responses to the target stimulus than monolinguals. Finally, the behavioral and conflict-monitoring ERP responses correlated with language proficiency and usage scores. Together, these results show that when tasks demands are high and auditory processing is part of the task, bilingual adults outperform monolinguals due to better stimulus identification and greater efficiency in managing task demands.

Keywords: bilingualism; executive control; neuroplasticity; selective attention; auditory processing; conflict monitoring

1. Introduction

The human brain is highly adaptable and has the capacity to change, remodel and reorganize in response to experiences across the lifespan (Pascual-Leone, Amedi, Fregni, & Marabet, 2005). A bilingual upbringing is suggested to be one example in which experience-related neuroplasticity (Fuchs & Flügge, 2014) leads to a slightly different neural organization as compared to a monolingual one (Calabria, Costa, Green, & Abutalebi, 2018). A debate spanning decades is the whether or not these changes extend beyond the classic perisylvian language system to the domain-general executive control network (Del Maschio & Abutalebi, 2018; Diamond, 2013). Most evidence in favor of this position follows from developmental studies in which bilingual children outperform monolingual peers in executive control tasks (Bialystok, 2017; Peal & Lambert, 1962; but see Lehtonen et al. 2018).

Both languages are constantly active and available in the bilingual mind, even in monolingual contexts (Kroll, Dussias, Bogulski, & Kroff, 2012; Thierry & Wu, 2007). Evidence for this notion comes from lexical-decision tasks in which a bilingual's performance is (a) facilitated by cognates, i.e. words which share both (similar) orthography and semantics in either language (e.g. *apple*; *appel* in Dutch), yet (b) impeded by interlingual homographs (or false friends), i.e. words which share orthography but not semantics (e.g. *"room" in Dutch translates to "cream"*; van Heuven & Dijkstra, 2010). Further evidence for language co-activation comes from Thierry and Wu (2007) who asked Chinese-English bilinguals to judge whether English word pairs were semantically related or not. Half of the written word pairs had a common character when translated into Chinese. Event-related brain potentials (ERPs) showed that despite the semantic-judgement task only requiring their second language (L2), bilinguals unconsciously translated the English words into their first language (L1; Chinese). Similar results have been found for American Sign Language – English bilinguals (Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011), suggesting that regardless of language combinations, bilinguals implicitly activate both their languages in monolingual contexts.

To achieve fluent linguistic performance, it is assumed bilinguals need to resolve continuous conflict between their languages (cf. Adaptive control model, Green & Abutalebi, 2013; Executive attention model, Bialystok, 2017) and as a consequence, a bilingual's executive control capacity is assumed to increase, including mechanisms such as inhibition, selective attention, working memory and conflict monitoring (Abutalebi & Green, 2007). Behavioral evidence for these proposals typically involves tasks such as the Flanker task (Eriksen & Eriksen, 1974), the Simon task (Simon, 1969), and versions of the Stroop task (MacLeod & MacDonald, 2000), which require some form of executive control engagement to resolve an element of conflict. In these tasks, the *bilingual advantage* refers to two main effects: Firstly, bilinguals show less interference between congruent and incongruent trials than monolinguals; Secondly, bilinguals display shorter reaction times (RTs) for conflict resolution than monolinguals (Costa, Hernández, Costa-Faidella & Sebastián-Gallés, 2009; Barac, Bialystok, Castro, & Sanchez, 2014; Bialystok, Craik, Klein, & Viswanathan, 2004). However, young adults often do not show any bilingual advantage (Bialystok, Martin, & Viswanathan, 2005; Gathercole et al., 2014; Kousaie & Phillips, 2012; Paap & Greenberg, 2013; Paap & Sawi, 2014; Prior & Gollan, 2013; Scaltritti, Peressotti, & Miozzo, 2017; von Bastian, Souza, & Gade, 2016), which has led to the suggestion that the bilingual executive control advantage has benefited from publication bias (de Bruin, Treccani, & Della Sala, 2015) and the use of unreliable and invalid measures (Paap, Johnson, & Sawi, 2015). For example, in a recent study by Paap and colleagues (2020), a large cohort of young adults completed language history questionnaires, intelligence tests, personality tests, a demographic questionnaire, as well as a series of typical executive control tasks. The results showed that fluid intelligence and sex predicted performance in the executive control tasks whereas other variables (e.g., bilingual status, SES, and personality) did not.

In response, proponents of the bilingual advantage argue that null results of the young adult population are due to this age group being at the height of their cognitive capacity

(Bialystok, 2016; 2017). Therefore, simple conflict tasks that show a bilingual advantage in children result in ceiling performance in young adults and thus no difference between language groups (Bialystok, 2006). This notion was supported by Costa and colleagues (2009) who did not find language group difference on a predictable low-monitoring version of a flanker task whereas they did in a high-monitoring version of the task. Similar results were found on a demanding conjunction search task but not on a simple visual feature search task (Friesen, Latman, Calvo, & Bialystok, 2014; but see Paap et al. 2018). Further, a large-scale study that controlled for various demographic factors between monolingual and bilingual language groups did not find any effects of bilingualism on typical executive control tasks (e.g., a Simon task, flanker task, and Stroop task; Anton, Carreiras, & Duñabeitia, 2019). However, they did observe better performance by bilinguals in a working-memory task, which suggests that bilingualism may not increase all aspects of executive control to the same extent. Other arguments for reported null findings or failures to replicate a bilingual advantage in specific tasks are that participant selection and the classification as bilingual are not the same across studies and the linguistic component of the task is often not the same (e.g., Grundy & Bialystok , 2019). Moreover, absence of a difference in behavioral differences does not imply the cognitive processes involved in performing the task are the same.

Neuroimaging studies have revealed a network of cortico-subcortical brain structures implicated in language control that is tightly bound to the domain-general executive control network in which the anterior cingulate cortex (ACC) plays a major role (Abutalebi & Green, 2016; Baumgart & Billick, 2017; De Baene, Duyck, Brass, & Carreiras, 2015). The ACC is not typically involved in monolingual language processing (e.g. Friederici, 2011), since its primary function lies in executive control processes such as conflict monitoring, error detection and response control (Bush, Luu, & Posner, 2000; Carter & van Veen, 2007). However, bilinguals recruit the ACC during language processing, including translation, language switching, and cross-language conflict resolution, indicating a bilingual co-dependence of language and executive processing (Abutalebi et al., 2007; Guo, Liu, Misra, & Kroll, 2011; Price, Green, & Von Studnitz, 1999; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008). As a result, it is suggested that bilingual adults appear to recruit the ACC more efficiently than matched monolinguals and have increased ACC grey matter volume (Abutalebi et al., 2012). These and further neuroplastic changes in bilingual grey and white matter (for a review see Li , Legault, and Litcofsky, 2014) are thought to be the cause of a general behavioral advantage on cognitive control tasks throughout the lifespan (Del Maschio & Abutalebi, 2018).

The electrophysiological signature of ACC activation are variants of the N2 component and the error-related negativity (ERN; Folstein & Van Petten, 2008). The frontocentral N2 is a negative-going waveform which peaks 200-350ms after a stimulus is presented and is thought to reflect early conflict detection (Folstein & Van Petten, 2008). In high conflict monitoring tasks, young adult bilinguals and L2 learners have been found to elicit larger N2 amplitudes and shorter N2 latencies than age-matched monolinguals, suggesting bilinguals allocate more neural resources to conflict monitoring than monolinguals (Fernandez, Tartar, Padron, & Acosta, 2013; Moreno, Wodniecka, Tays, Alain, & Bialystok, 2014; Morales, Yudes, Gómez-Ariza, & Bajo, 2015). However, causal links between these neutral responses and behavior are rarely found, and it is therefore unclear how greater executive control is expressed in different behavior i.e. it is unknown what bilinguals do differently than monolinguals that results in differences in behavioral measures.

1.2 The current study

We modified the arrow flanker task (Eriksen & Eriksen, 1974) in two ways to increase task difficulty, which as explained above, may lead to behavioral differences between monolingual and bilingual language groups. In the standard flanker task participants respond following the direction of the central arrow (henceforth “Same” task) and we added an “Opposite” task in which participants are required to respond to the opposite direction of the central arrow. Participants did not know which task was to be performed until hearing a cue (a high tone or a low tone). Our second manipulation was to vary task preparation time with three different Stimulus Onset Asynchronies of the task cue and target stimulus (SOA: -1s, -500ms, 0ms). We predicted that language group differences in RTs and error rates would gradually emerge with increasing task demands and be established in the most demanding experimental condition, the Opposite task with 0ms SOA. Besides analysis of behavioral measures upon which many of the reports on the “bilingual advantage” are based, we also analyzed cue, target, and response ERPs to compare task preparation, target conflict-monitoring, and response monitoring processes, respectively. The ERP components of interest here are the N2 due to its sensitivity to conflict monitoring (Folstein & Van Petten, 2008) and the fact that this component has previously been found to be affected by language status. Further, we anticipated to find language group differences in the (auditory) P2, which in previous studies has been found to be greater in bilinguals (e.g., Kuipers & Thierry 2012). The Correct Response Negativity (CRN; Bartholow, Pearson, Dickter, Sher, Fabiani, & Gratton, 2005) was also analyzed due to its sensitivity to response monitoring. With these measures we can track performance in the three different stages of the task: preparation, target processing, and response monitoring. Finally, to establish if any of these measures is responsible for the behavioral pattern of results, we correlated the above measures with self-reported language proficiency and usage measures of both experiments reported below. Based on previous studies mentioned above, we expected to find that bilinguals would display greater N2 responses, however, given that it is unknown if bilinguals use a different strategy to perform the task, we could not predict at which of the cue and target stimuli this language effect would appear. Finally, the CRN would complement the N2 analysis because any language group differences at earlier processing stages may be reversed at later ones.

2. Materials and methods Experiment 1

2.1 Participants

Thirty-six participants (26 female; 4 left-handed, Oldfield, 1971), with ages ranging from 18 to 25 years (*M* = 19.7, *SD* = 1.94) were recruited from the University of Stirling. All participants were asked to provide language-proficiency and language-usage ratings about English as well as their other language(s), if applicable, using the most commonly used questions is such a questionnaires (Li, Sepanski, & Zhao, 2006). Participants were classified as either monolingual or bilingual based on these language proficiency measures. Written informed consent was obtained from all participants. The study was approved by the Ethics Committee of the University of Stirling and testing was carried out in accordance of the Declaration of Helsinki.

2.2 Language groups

Participants were split into 18 monolinguals and 18 bilinguals who neither differed in age ($t(34) = .511$, $p = .613$) nor on gender ($X^2(1) = .554$, $p = .457$). Monolinguals were native English speakers who had either lived in Scotland all their lives (11) or moved there from England, Northern Ireland, Wales or the US for their studies or at an earlier stage in their life. With an average acquisition age of 12.1 years (*SD* = 1.64), 14 reported to have previously studied a language, but to speak English 100% of the time (*M* = 99.72%, *SD* = 1.18%). No participant in the monolingual group regarded themselves proficient in any other language other than English.

Two thirds of bilinguals were born in a non-English-speaking country and only two in Scotland. However, all bilinguals reported to speak English fluently, with total scores for proficiency (all subcomponents pooled together) ranging from 6 to 10 and a median of 10. Bilinguals stated they use English 63.61% of the time (*SD* = 12.10) and the other language

36.39% (*SD* =12.10) of the time. All bilinguals (except for one Catalan – Spanish bilingual) spoke English as either their L1 or L2. Therefore, proficiency and usage scores given to English and the other language were allocated to L1 (first acquired and dominant language) or L2 (secondly ranked in acquisition and dominance), dependent on each individual's reported age of acquisition and dominance. Details of the Catalan-Spanish individual's Spanish (L2) proficiency are not reported since the questionnaire only focused on details of English and one other language. Bilinguals' L1 included one of the following languages: English (8), French (1), Italian (1), Slovene (1), Czech (1), German (3), Catalan (1), Slovak (1) or Norwegian (1). Bilinguals' L2 included either English (9), Dutch (1), Italian (1), Norwegian (1), German (1), Urdu (2), Spanish (1), Burmese (1) or Swedish (1). L2 acquisition was reported from birth (4 participants) to 9 years with a mean age of 4.11 (*SD* = 3.13). All bilingual participants were highly proficient in both languages, with a total median score of 10 (range of 7 to 10) for L1, and a total median score of 9 (range of 1 to 10) for L2. Wilcoxon signed rank tests showed that bilinguals rated their L1 proficiency higher than their L2 proficiency: Reading ability, $Z = -2.848, p = .004$; Speech comprehension, $Z = -2.399, p = .016$; Conversational fluency, $Z = -2.406, p = .016$; Writing proficiency, $Z = -2.238, p = .025$.

Mann-Whitney *U*-tests were conducted to compare the two language groups on their L1 and L2 proficiency and usage scores (see Table 1). Unsurprisingly, monolinguals and bilinguals were highly proficient in their L1 and the two groups did not differ on any of the four L1 proficiency subcomponents (all p's > .05). Regarding L1 usage, the groups significantly differed on all subcomponents (all p's < .05). This may be expected since bilinguals allocate part of their time to their L2. Crucial for the differentiation and thus definition of the two language groups, monolinguals and bilinguals significantly differed on their L2 proficiency and usage scores ($p < .01$). For reasons mentioned above, we did not calculate language dominance from these measures, but our sample reflects the broad range of L2 proficiency found in society ranging from simultaneous bilinguals to relatively late learners. Also, one could argue there may be an immigrant status and/or SES confounds in our data since far more bilinguals than monolinguals were from Scotland. However, RT differences in flanker tasks have been found in bilingual groups that had not emigrated (Costa et al., 2009) and as mentioned above, SES does not appear to contribute to contribute to RT variance in typical executive control tasks (Paap et al., 2020).

Table 1. L 1 and L2 proficiency and usage of participants of Experiment 1.

	Monolinguals (N = 18)		Bilinguals (N = 18)		Mann-Whitney test	
L1 Proficiency	Median	Range	Median	Range	U	p-value
Reading Ability[a]	10	8 - 10	10	7 - 10	153	0.706
WritingAbility[a]	10	8 - 10	10	7 - 10	119	0.103
Conversational Fluency[b]	10	9 - 10	10	7 - 10	142	0.329
Speech Comprehension[C]	10	9 - 10	10	8 - 10	143	0.283
L2 Proficiency						
Reading Ability[a]	1	1 - 4	8	1 - 10	10.5	0.000
WritingAbility[a]	1	1 - 5	8	1 - 10	12.0	0.000
Conversational Fluency[b]	1	1 - 4	9	6 - 10	0.0	0.000
Speech Comprehension[C]	1	1 - 4	10	5 - 10	0.0	0.000
L1 Usage						
Reading Time[d]	10	7 - 10	8.5	5 - 10	81.5	0.000
Writing Time[d]	10	10	8.0	4 - 10	63.0	0.000
Speaking Time[d]	10	10	8.5	5 - 10	36.0	0.000
Media Time[d]	10	7 - 10	8.5	3 - 10	78.5	0.000
L2 Usage						
Reading Time[d]	1	1 - 2	8	1 - 10	11.0	0.000
Writing Time[d]	1	1 - 2	8	1 - 10	9.5	0.000
Speaking Time[d]	1	1 - 3	8	3 - 10	.5	0.000
Media Time[d]	1	1 - 2	8	1 - 10	19.0	0.000

Notes. Self-report on a scale of 1-10:
a: 1 = not literate; 10 = very literate
b 1 = not fluent; 10= very fluent
C 1 = unable to understand conversation; 10 = perfectly able to understand conversation
d 1 = not at all; 5 = sometimes;10 = always

2.3. Materials

The visual stimuli were 5 white arrows which were centrally aligned on a black background (visual angle 11 degrees). The 4 flanker arrows were either directed in the same or opposite direction as the central target. The visual stimuli were paired with one of two auditory stimuli, a high tone 1200Hz or a relatively low tone 1000Hz presented at 70dB, which indicated whether the Same or Opposite task should to be performed. The tone-task mapping was counterbalanced across participants.

2.4. Procedure

Stimuli were presented using E-prime software (Psychology Software Tools, Inc.) on a 19 inch monitor with a 60Hz refresh rate placed at a distance of approx. 80 cm from the participant. Auditory stimuli were presented via two speakers placed next to the monitor. Behavioral responses were recorded with a 5-button Response Box.

Participants received written instructions about the tasks and the tone cue before receiving 30 practice trials to familiarize themselves with the different experimental conditions and the tone-task mapping. They were encouraged to make speeded and accurate decisions using the outer left and right buttons on the response box with their left and right index fingers. The experimenter remained in the testing room to ensure the participant understood the task. Each experimental trial began with a fixation cross for 700ms, followed by the tone cue (200ms) with varying SOAs (-1s, -500ms, 0ms) relative to the onset of the target stimulus which was presented until the participant's response or for a maximum duration of 1500ms (see Figure 1). A total of 576 trials of all 12 conditions (48 trials per condition) were presented in a randomized order, meaning participants did not know which task to perform until they heard the tone. Participants received short breaks every 100 trials.

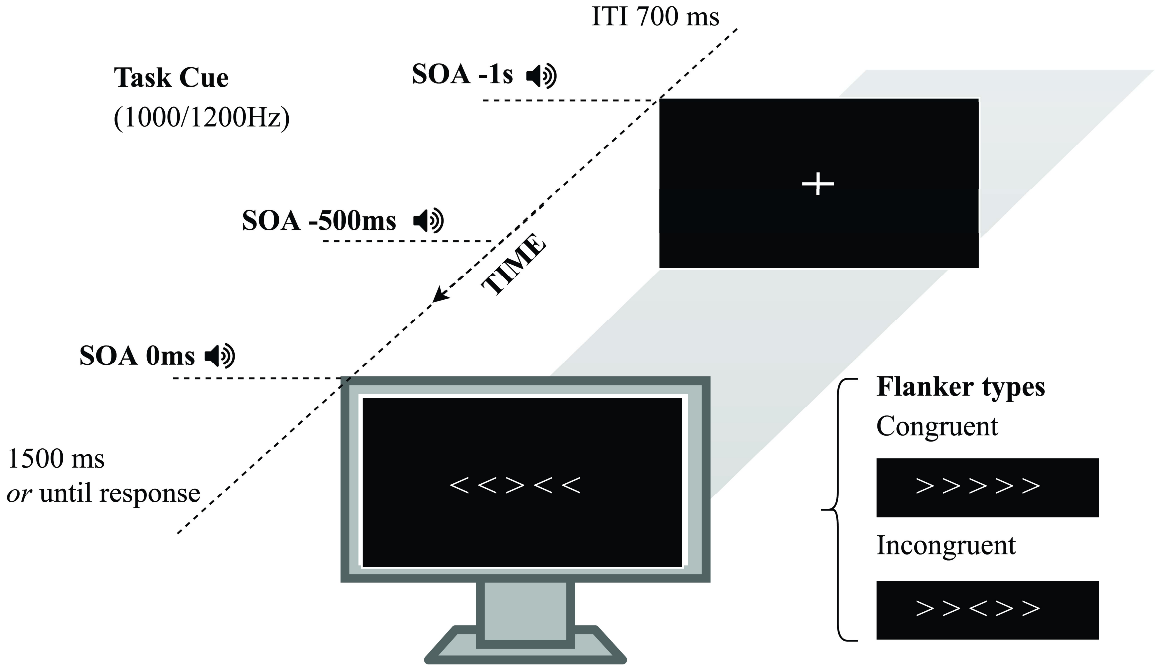

Figure 1. Trial procedure and experimental conditions.

2.5. ERP Data Acquisition

Electrophysiological data was continuously recorded by using a SynAmps II amplifier system connected to a 64-channel Ag/Ag/Cl Quik-Cap (extended International 10-20 system; NeuroMedical Supplies). Eye movements were recorded with a horizontal electro-oculogram (HEOG) and vertical electro-oculogram (VEOG) using electrodes placed on the lateral canthi of each eye, as well as below and above of the left orbit, respectively. Two additional electrodes were placed over the left and right mastoids. Data was recorded at a rate of 1kHz and filtered between 0.1Hz and 200Hz. The reference electrode was located between Cz and Cpz and the ground electrode was located between Fz and Fpz. Electrode impedances were kept below 5kΩ. Offline EEG data were Low-pass filtered at 30Hz (48db/Oct) and re-referenced to the average of the mastoid electrodes (Picton et al, 2000). Eye blink artefacts were mathematically corrected (Gratton, Coles, & Donchin, 1983). Target stimuli epochs ranged -100 to 900ms relative to stimulus onset whilst cue and response-locked epochs ranged from -100 to 500 ms relative to their onset. Epochs containing artefacts exceeding ±60μV on any electrode were rejected.

2.6. Analyses

Reaction times were analyzed with linear mixed effects modelling using lme4 package (Bates, Maechler & Bolker, 2012) implemented in R (R Core Team, 2012). We started with the most complex model and removed factors one-by-one whilst obtaining p-values using likelihood ratio tests comparing the more complex model against a simpler one (following Winter, 2013). Further p-values for significant main effects and estimated marginal means of interactions were obtained using lmerTest (Kuznetsova, 2019). RTs below 200ms were excluded from analysis (2.4% for monolinguals 1.57% for bilinguals). Post-hoc analyses following significant interactions focused on the critical differences driving the interaction in question. Errors were analyzed in a similar way albeit using logistic linear modelling as implemented in the lme4 package.

Analysis of ERPs followed visual inspection of grand average waveforms of predicted ERP components of interest, elicited by the cue, the target, and manual responses. The cue (at both negative SOAs) elicited an N2 conflict-monitoring response task which appeared to differ between language groups (Figure 2). Mean amplitude of the N2 was calculated between over 240-300ms of frontocentral electrodes (F1, Fz, F2, FC1, FC2, FCz) in accordance with the N2-typical time-window and topography (Folstein & Van Petten, 2008).

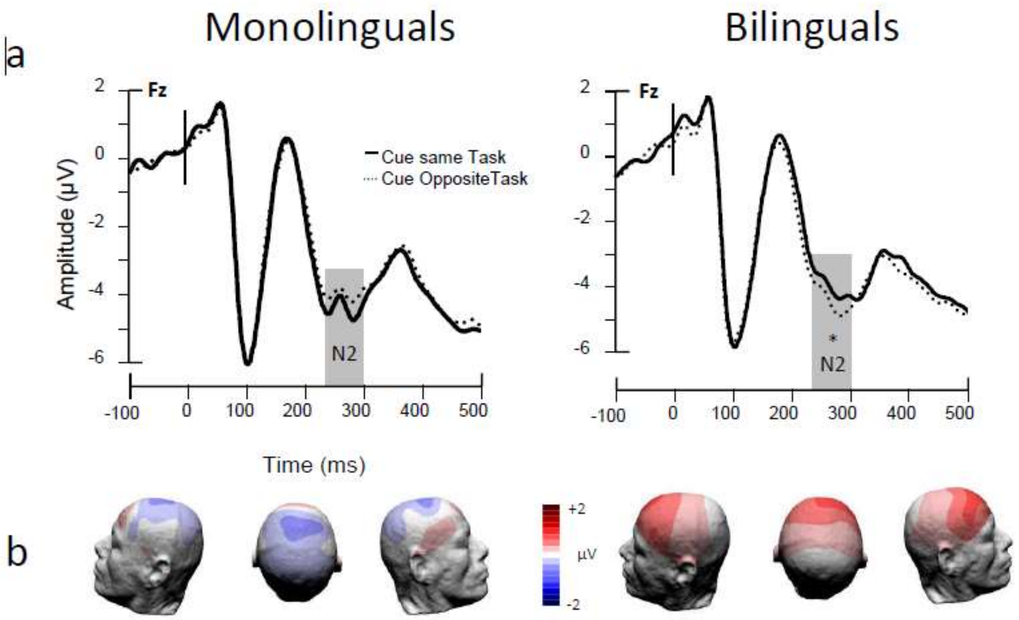

Figure 2. Grand average ERP waveforms of the cue stimulus.

a. Cue-elicited grand average ERP waveforms of representative electrode (Fz) for each task and group. The grey-shaded area indicates the N2 time-window used for analysis and the * indicates significance of the difference (p<.05). Stimulus onset is indicated by the vertical black line at 0ms.

b. Topography of the task difference for each group.

ERP waveforms of the target stimulus show clear P2 and N2 modulations at frontocentral electrodes (Figure 3). However, the P2 peaked earlier in the 0ms than in the -500ms and -1s SOA conditions, which is most likely due to the fact that at 0ms the stimulus was a compound of both an auditory and a visual stimulus whereas at the -500ms and -1s SOAs the stimulus was visual only (Key, Dove, & Maguire, 2005). Therefore, based on the pattern visible in the grand average waveforms, different P2 time-windows were chosen for 0ms SOA (170-240ms) and the -500ms and -1s SOAs (200-250ms) to capture the time-windows of maximal modulation at each SOA.

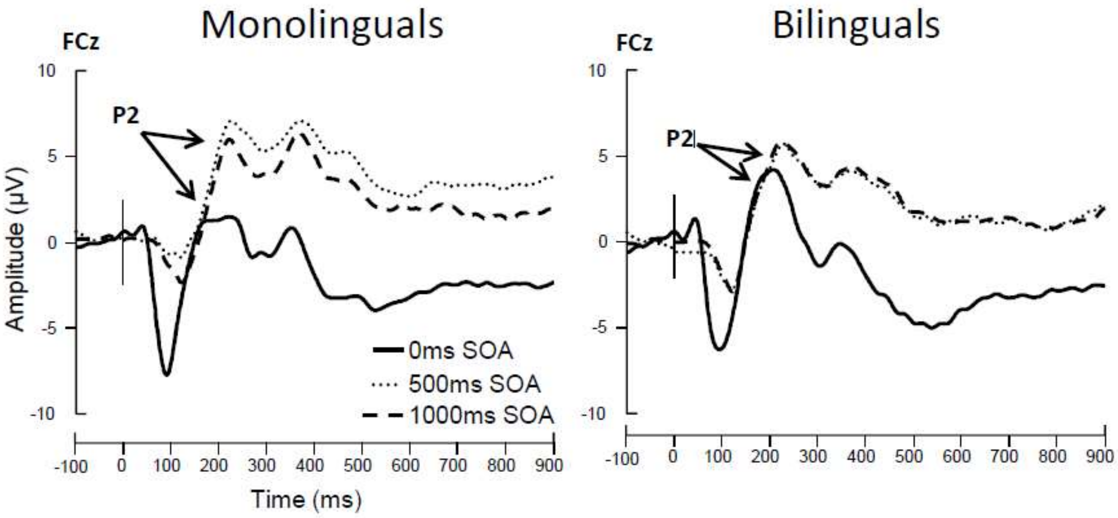

Figure 3. Grand average ERP waveforms of the target stimulus.
Grand average ERP waveforms of representative electrode FCz of each SOA for each group. The arrows indicate the P2 which clearly differs in the visual alone stimulus at SOA -500 and -1s differs from the combined auditory and visual one at 0ms SOA. Stimulus onset is indicated by the vertical black line at 0ms.

Inspection of the P2 at 0ms SOA revealed a striking difference between groups (Figure 4). The monolingual P2 is collapsed at this SOA whereas it is mostly preserved in bilinguals. The topography of the P2 shows a typical central-frontocentral distribution (Figure 4b; Crowley & Colrain, 2004), which is why electrodes FC1, FCz, FC2, C1, Cz, and C2 were included in the analyses.

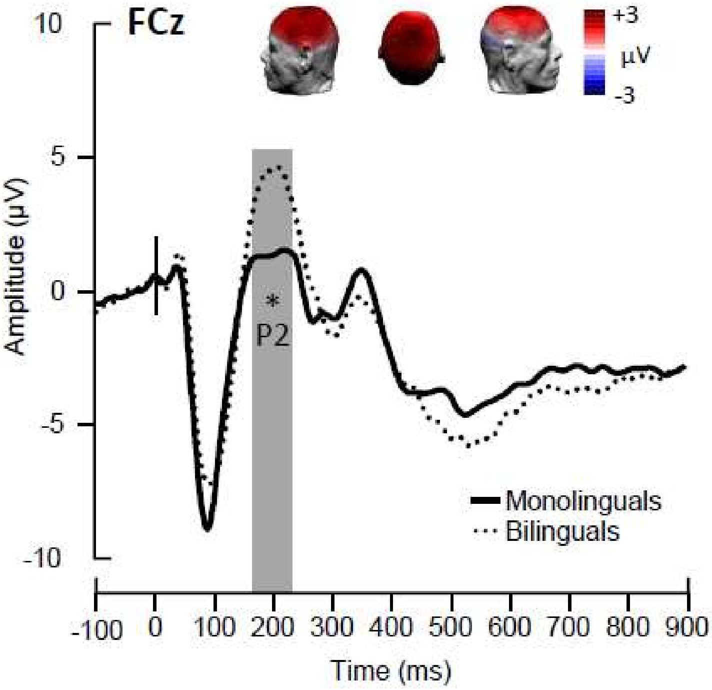

Figure 4. Grand average ERP waveforms and topography of the target P2. Target-elicited grand average ERP waveforms of representative electrode (FCz) shows the overall group difference (monolinguals black line, bilinguals red line) at 0ms SOA and the topography of the language group difference (bilinguals –monolinguals). The grey-shaded area was used for the analysis of mean P2 amplitude. Stimulus onset is indicated by the vertical black line at 0ms.

The congruency N2 elicited by the target stimulus was maximal over frontocentral electrodes (F1, Fz, F2, FC1, FC2, FCz) and analyzed from 290-310ms (Fig 5).

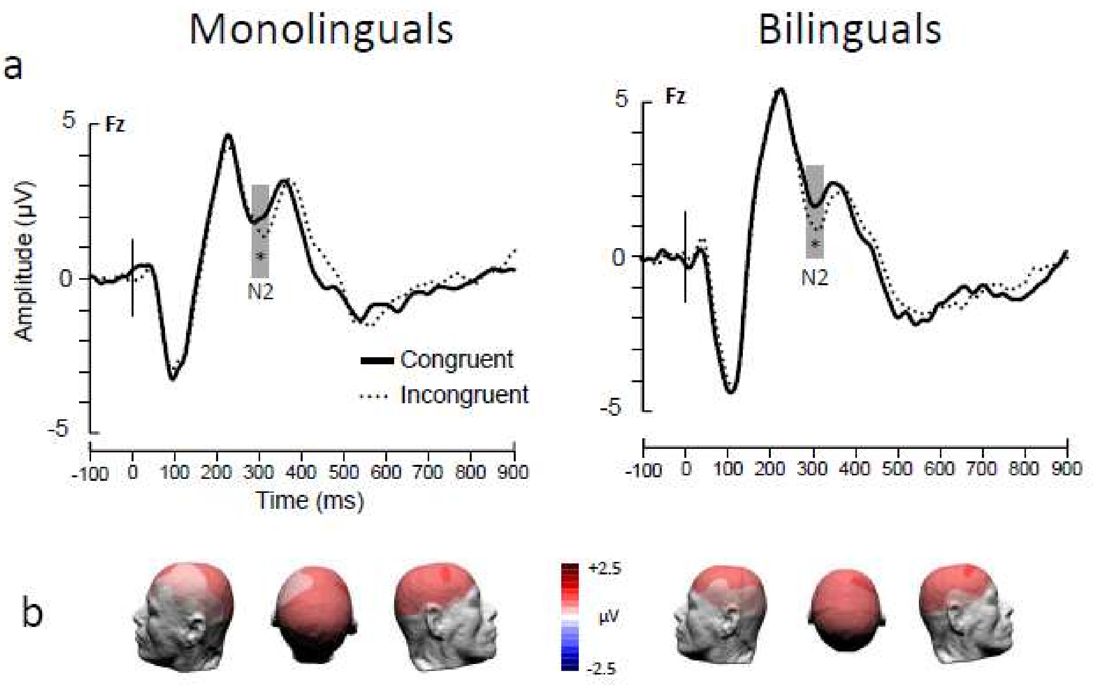

Figure 5.

a. Target-elicited grand average ERP waveforms of representative electrode (Fz) for Congruent (black) and incongruent (red) conditions and each group. The grey-shaded area was indicates the N2 time-window used for analysis and the * indicates significance of the difference (p<.05).

b. Topography of the N2 congruency effect (congruent – incongruent) for each group.

To gain insight in response monitoring processes we also analyzed response-locked ERPs. Given that there were too few errors for a meaningful ERN analysis, we measured the Correct Response Negativity (CRN) which also indexes performance monitoring (Bartholow, Pearson, Dickter, Sher, Fabiani, & Gratton, 2005; Vidal, Hasbroucq, Grapperon, & Bonnet 2000). There was a clearly visible CRN response from 20-60ms over frontal electrodes, in line with the well documented ERN and CRN (Bartholow et al., 2005; Falkenstein, Hoorman, Christ, & Hohnsbein, 2000; Yeung, Botvinick, & Cohen, 2004; Figure 6). We therefore analyzed mean CRN amplitude in this time-window over frontal electrodes (FP1, FP2, FPz, AF3, AF4, F1, F2, Fz) where its amplitude is typically maximal.

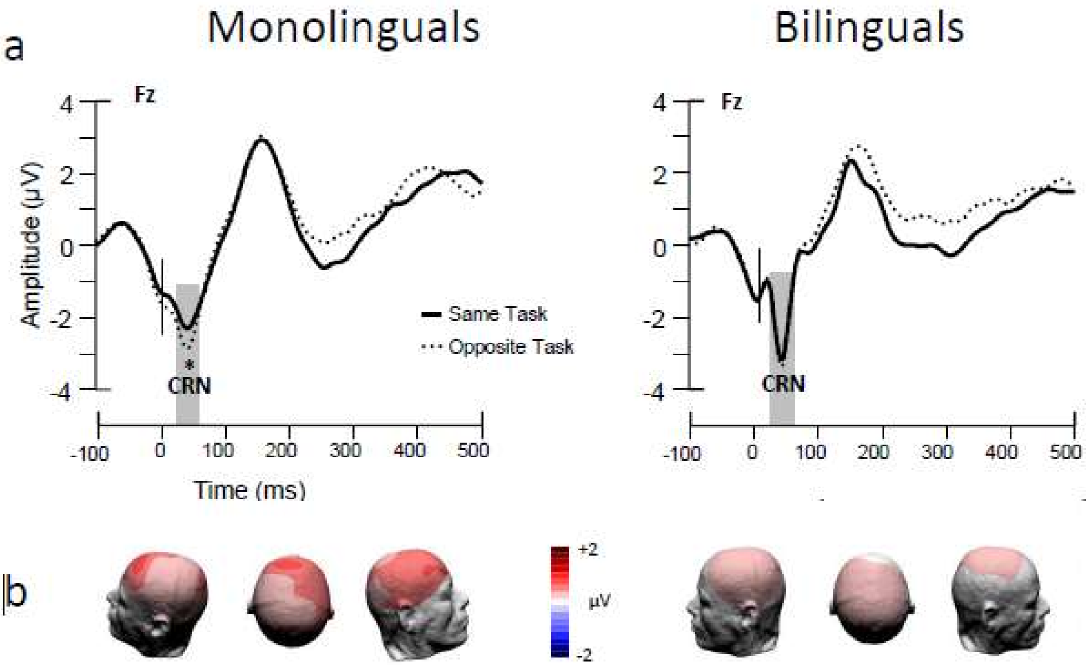

Figure 6.

a. Response-locked grand average ERP waveforms of representative electrode (Fz) for each task and group. The vertical bar at 0ms indicates response onset. The grey-shaded area was indicates the CRN time-window used for analysis and the * indicates significance of the difference (p<.05).

b. Topography of the task difference for each group.

Mean ERP amplitudes were analyzed using ANOVAs with Congruency (congruent vs., incongruent); Task (same vs. opposite), and stimulus-onset-asynchrony (SOA; -1s, -500ms, 0ms) as within-subjects factors and language group as a between-subjects variable. Greenhouse-Geisser corrections are reported where applicable.

3. Results

3.1. Behavioral data

Linear models on RTs (Table 2) with the 4-way and 3-way interactions did not converge, which is why we focus on the 2-way interactions with group. Task, Congruency, SOA, and their interactions with Group were entered as fixed effects with random intercepts for participants and random slopes for Task and Congruency only, for convergence reasons. In the final model on RTs, the main factor Task was significant (β = -17.86, t(32) = -2.58, p=.015) with shorter RTs on average in the Same task (654ms) than the Opposite task (683ms). Congruency was significant (β = 21.38, t(35)=6.02, p<.0001) showing a typical congruency effect (congruent: 658ms; incongruent: 679ms). The factor SOA was significant with longer RTs in the 0ms SOA (775ms) than the 500ms SOA (646ms; β = -129, t(18045) = 34.6, p<.0001), which in turn, had longer RTs than the 1s SOA (595ms; β = -180, t(18043) = 48.4, p<.0001). Importantly, the interaction between Task and Group was also significant (β = -24.92, t(32)=-2.49, p=.017), with bilinguals displaying a smaller, but still significant, RT difference between tasks (Opposite task 18ms slower than the Same task, p=.04) than monolinguals (a difference of 43ms in the same direction, p<.0001).

In the logistic model on errors the best fitting model only included the factors Task and Group. The factor Task was significant (β = .29, z = 5.24, p < .0001) showing that, overall, fewer errors were made in the Same task (6.2%) than in the Opposite task (8.14%). The factor Group was also significant (β = -.17, z = -3.19, p = .001) revealing that, overall, bilinguals produced fewer errors (6.6%) than monolinguals (7.8%).

Table 2. RTs and standard deviation (SD) for each task (Same; Opposite), SOA (-1000; -500; 0) and congruency (Congruent (con.); Incongruent (Incon.)), and the congruency effect (diff) of Experiment 1.

	Same task								
	-1000			-500			0		
	Con.	Incon.	diff.	Con.	Incon.	diff.	Con.	Incon.	diff.
Bil.	558 (102)	583 (106)	25	606 (107)	640 (103)	34	756 (111)	786 (105)	30
Mon.	566 (95)	602 (89)	36	612 (79)	634 (82)	22	743 (87)	767 (82)	23
	Opposite task								
	-1000			-500			0		
	Con.	Incon.	diff	Con.	Incon.	diff.	Con.	Incon.	diff.
Bil.	578 (114)	601 (103)	23	629 (109)	658 (110)	29	768 (101)	770 (100)	2
Mon.	619 (100)	630 (87)	1	689 (97)	689 (112)	0	794 (95)	795 (100)	1

Note. Bil. = Bilinguals Mon. = Monolinguals

To examine whether the above group differences are reflecting group differences in task preparation strategies, target stimulus processing, or performance monitoring, we next analyzed the ERPs elicited by the task cues, target stimulus, and the response. For the cue analysis, data of the 500ms and 1s SOAs were averaged, because when perceiving the cue, the participant does not yet know the SOA to the target stimulus.

3.2. Cue elicited ERPs

The analysis of mean N2 amplitudes revealed a significant Task by Group interaction F(1,33) = 4.0, p = .03) only. Follow-up analyses for each group showed that the effect of task was significant in bilinguals (F(1,17) = 4.8, p =.04) with the Opposite task (-2.9µV) eliciting a more negative N2 than the Same task (-2.6µV). There was no significant effect of Task in monolinguals (p > .2). Therefore, the cue for the Opposite task alone elicits a greater conflict monitoring response than the Same task cue in bilinguals, which suggests bilinguals have different task preparation strategies than monolinguals.

3.3. Target -elicited ERPs

The ANOVA on target-elicited mean P2 amplitudes revealed significant main effects of SOA (F(1.4,45) = 18, p < .001) showing that the 0ms SOA elicited a smaller P2 (2.6µV, SD = 3.0) than the 500ms SOA (5.1µV, SD = 3.7) and 1s SOA (5.2µV, SD = 3.0). The interaction between SOA and Group was also significant (F(2,66) = 6.8, p = .002). Post-hoc t-tests on the groups' P2 at each SOA showed that the P2 group difference of 2.2µV at 0ms was significant (t(1,16) = 2.9, p = .011) whereas this was not the case for 500ms (0.8 µV difference; p > .3) and 1s (0.3µV difference; p > .6). Therefore, the bilinguals display a larger P2 than monolinguals when the tone, cueing which task needs to be performed, and the arrow stimuli are presented at the same time.

The analysis on target-elicited mean N2 amplitudes revealed significant main effects of Congruency (F(1,33) =9.5, p = .004), showing the typical more negative peak for incongruent (2.0µV) than congruent (2.6µV) trials, and SOA (F(2,66) = 46, p < .001) with the largest (most negative) N2 for 0ms (-.9µV, followed by 1s (3.1µV), and 500ms (4.6µV; all comparisons p < 0.01). The interaction between Congruency and SOA was also significant (F(2,66) = 4.5, p < .015), with the congruency effect not significant at 0ms (<.001 µV difference, p = .99) whereas it was significant at 500ms (.87µV difference, p = .001) and at 1s (.87µV difference, p=.009).

3.4. Response locked ERPs

The analysis on CRN amplitude revealed a significant main effect of SOA (F(2,66) = 26. p < .001), with the 0ms SOA (-1.0µV, SD = 1.7) eliciting a less negative CRN than the -500ms SOA (-1.8µV SD 1.6), and -1s SOA (-2.1µV SD 1.6; both p<.001) whilst the latter two SOAs did not differ (p = .14). The Task X Group interaction was also significant (F1,33 = 5.6, p = .024). Follow-up analyses for each group separately revealed that the effect of Task was marginally significant in monolinguals (F(1,16) = 4.0, p = .06) with the Opposite task (-1.8µV SD 1.4) eliciting a more negative CRN than the Same task (-1.4µV SD 1.2), whereas there was no task difference in bilinguals (p =.6; both tasks -1.8µV SD 1.8). Finally, the interaction between Task, Congruency, and SOA was also significant (F(2,66) = 3.4, p = .038; see Figure 7) which, due to the absence of any group effects, was not analyzed and discussed further.

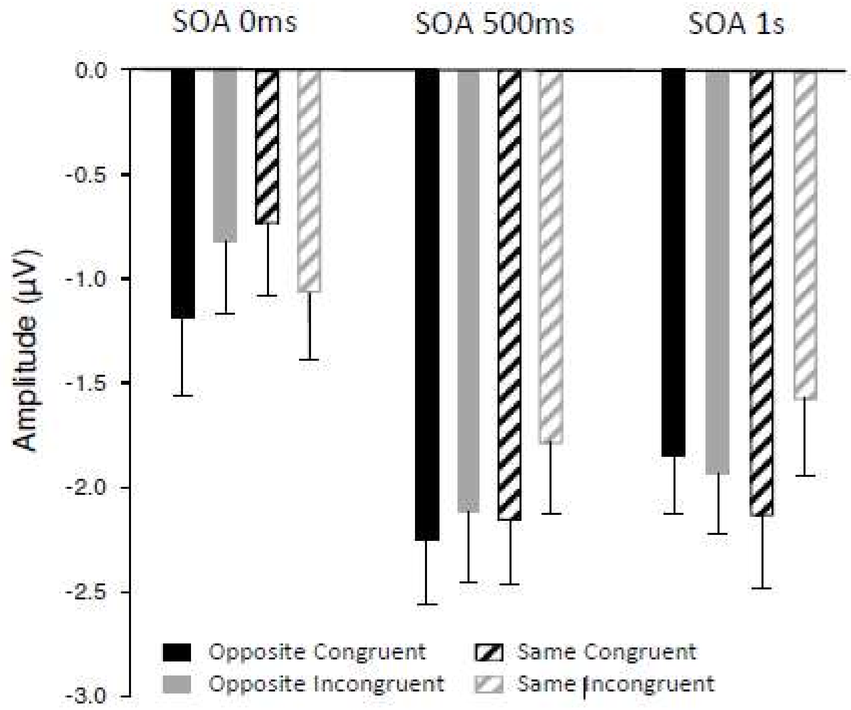

Figure 7. CRN amplitudes for each of the conditions included in the significant Task X SOA X Congruency interaction.

4. Discussion

To test the task-load hypothesis and identify potential stimulus processing differences between bilinguals and monolinguals, we recorded ERPs while participants performed low and high difficulty versions of the arrow flanker task. Further difficulty was created by shortening task preparation time. As predicted, bilinguals were less hampered by the more difficult task than monolinguals and displayed fewer errors overall suggesting they are more flexible in applying different task sets. However, the absence of a group interaction with SOA shows task preparation time does not appear to be an important factor determining task difficulty.

The ERP results shed light on the possible reason the presence, and absence of language group differences. The way participants prepare for each task is reflected in the ERP response to the task cue. The cue N2 conflict-monitoring response in bilinguals was greater in the Opposite than Same task, whereas monolinguals' cue N2 did not distinguish between tasks. Therefore, the bilinguals displayed a greater conflict monitoring response in preparation for the more difficult task.

Further, the P2 elicited by the target stimulus was found to be larger for bilinguals than monolinguals in the 0ms SOA condition only. There are several possible mechanisms that could underlie this result. Given the frontocentral topography of this P2 (Figure 3), it may be an auditory-driven P2 generated by primary and secondary auditory cortices (Zouridakis, Simos, & Papanicolaou, 1998) and associated with orienting of attention (Key et al., 2005; Morris, Steinmetzger, & Tøndering, 2016). If indeed due to attention allocation, the question is whether this P2 reflects top-down strategic attention allocation or stimulus-driven prioritization of auditory information.

During our multisensory task (at 0ms condition), strategic prioritization of the auditory cue may be the most appropriate and beneficial way of solving the task, since it is only presented for 200ms and indicates how to respond to the visual stimulus, which was presented for 1500ms. However, sensory competition may also result in modality dominance (Sandhu & Dyson, 2012), with the visual modality often dominating under high task demands (Posner, Nissen, & Klein, 1976; Sandhu & Dyson, 2012). Therefore, given that the auditory P2 decreases in amplitude when attention to the auditory signal of a multi-modal stimulus is reduced (Morris et al., 2016), the smaller P2 observed in monolinguals in our 0ms SOA condition could reflect that they prioritize processing the visual target.

The next question is why bilinguals did not prioritize the visual target stimulus. Chabal and colleagues (2015) showed bilinguals outperform monolinguals on an audio-visual object search task, which included conflicting, consistent, or neutral auditory cues. If auditory cues were automatically prioritized, conflicting auditory cues should have interfered with bilingual's performance, which was not the case. Therefore, visual tasks that include relevant auditory information may engage a bilingual’s executive and auditory system simultaneously; a possible mechanism which Krizman and colleagues (2014) refer to as ‘sensory-cognitive co-activation’ (p. 37). This co-activation may have been the catalyst for the increased attentional P2 response in bilinguals we observed.

4.2. The N2 and CRN reveal group differences in performance and conflict monitoring

The frontocentral target-elicited N2 displayed the typical greater negativity for the incongruent than congruent flanker stimuli. Contrary to some previous findings from go-nogo tasks (Fernandez, Acosta, Douglass, Doshi, & Tartar, 2014; Moreno et al., 2014), this conflict monitoring response (Folstein & Van Petten, 2008) did not differ between language groups. The failure to replicate the increased N2 in bilinguals may be due to a lack of robustness of the effect, or a carry-over from the increased positivity of the P2 at 0ms SOA to the N2. However, given that the congruency modulation of the target N2 was not qualitatively different when there was no increased P2 (at the 500ms and 1s SOAs), the latter interpretation is unlikely. Another possibility is that the greater N2 conflict-monitoring response to the cue removes the need for greater conflict monitoring of the target. A better preparation of the task makes target processing easier, which is in line with the better performance displayed by bilinguals.

Finally, the CRN was also differently modulated by language groups. The monolinguals displayed greater negativity in the more difficult Opposite task, whereas there was no task difference in bilinguals. Therefore, monolinguals appear to have a very late task monitoring process, i.e., after they had given their response, which bilinguals displayed much earlier during task preparation.

To investigate whether the above results are indeed due to bilingual prioritization of the auditory cue, we next repeated the experiment whilst using a visual task cue. To replace the auditory cue we chose to use a red and a blue square surrounding the stimuli whilst keeping the other variables the same.

5. Experiment 2

5.1. Participants

Forty participants were recruited using the same inclusion criteria as in Experiment 1 with no one having participated in Experiment 1. Five participants were excluded from analysis due to experiment abortion, not following the instructions, and/or too noisy EEG data. Of the 35 participants included in the analyses (aged 18 to 29; M = 21.34 years; SD = 3.21), 27 were female, and 17 were self-reported monolingual (leaving 18 bilinguals). The participants identified as Bilingual (mean age 22 years, SD=3.8) did neither differ in age from monolinguals (mean age 20 years SD=2.0; t(33)= 1.8, p=.076) nor on gender ($X^2(1) = .008$, $p = .9$). All monolingual participants reported high English proficiency (range 7-10, Median= 10, SD = .9) and a low proficiency in any other language (range 1-5; Median = 1, SD = 1.1). Two monolingual participants reported using a second language 5% of the time compared to the others who reported 0% and originated from native English-speaking countries (UK and US). Bilinguals' L1 included one of the following languages: English (2), Italian (5), Spanish (5), German (1), Polish (1), Bulgarian(1), Romanian (1), Gaelic (1), Portugese (1). Bilinguals' L2 included either English (16), French (1), and Spanish (1). Of these, 3 reported being simultaneous bilinguals and all started learning English between birth and 12 years (M = 5; SD = 5). Importantly, the language groups did not differ on any L1 proficiency measure, but they did differ significantly on all L2 measures (see table 3).

Table 3. *L 1 and L2 proficiency and usage of participants of Experiment 2.*

	Monolinguals (N = 17)		Bilinguals (N = 18)		Mann-Whitney test	
L1 Proficiency	Median	Range	Median	Range	U	p-value
Reading Ability[a]	10	5 - 10	10	8 - 10	149	0.909
WritingAbility[a]	10	6 - 10	10	7 - 10	146	0.832
Conversational	10	6 - 10	10	7 - 10	125	0.369

Fluency[b]						
Speech Comprehension[C]	10	8 - 10	10	8 - 10	138	0.636
L2 Proficiency						
Reading Ability[a]	1	1 - 7	8	2 - 10	4.5	0.000
WritingAbility[a]	1	1 - 5	8	2 - 10	4.5	0.000
Conversational Fluency[b]	1	1 - 4	9	4 - 10	0.5	0.000
Speech Comprehension[C]	1	1 - 4	8	7 - 10	0.0	0.000
L1 Usage						
Reading Time[d]	10	9 - 10	8	5 - 10	28	0.000
Writing Time[d]	10	10	8	3 - 10	34	0.000
Speaking Time[d]	10	10	9	6 - 10	42	0.000
Media Time[d]	10	8 - 10	8.5	2 - 10	46	0.000
L2 Usage						
Reading Time[d]	1	1 - 3	8	1 - 10	10	0.000
Writing Time[d]	1	1 - 2	7.5	4 - 10	0.0	0.000
Speaking Time[d]	1	1 - 2	8	4 - 10	0.0	0.000
Media Time[d]	1	1 - 3	7	1 - 10	10	0.000

5.2. Methods

The same materials and procedure was used as in Experiment 1 except that the task cues were a red and or a blue square (side length 66mm; line thickness 4mm) within which the fixation cross (at negative SOAs) or the target stimulus (at 0ms SOA) were presented. The cue remained on screen until the response deadline of 1500ms not to induce any stimulus offset ERPs.

5.3. Analysis

The data were analyzed the same way as in Experiment 1, using the same electrodes and time-windows for each component. The one exception to this is that due to the visual modality of the task cue, the early visual cue ERPs (P1 and N1) are located at occipital electrodes as compared to central one for the auditory cue used in Experiment 1 (Luck & Hillyard, 1994). Therefore, according to the typical topography and latency of the N1 (Vogel & Luck, 2000), its amplitude was analyzed over PO7 and PO8 from 150-210ms (Fig 8). For the same reasons, the P2 was analyzed from 190-240ms, the N2 from 300-340ms using the same electrodes as in Experiment 1.

6. Results

In the analysis on RTs (Table 4), systematic removal of factors from the most complex model led to one that included the factors Task, Congruency, their interaction, and SOA, as fixed factors with random intercepts for participants, and random slopes for Task and Congruency only for convergence reasons. The factor Task was significant (β = -28.0, t(59) = -5.1, p<.0001) resulting from overall shorter RTs in the same task (695ms) than in the opposite task (711ms). The factor Congruency was significant (β = 32.7, t(57)=5.4, p<.0001) with shorter RTs in the congruent condition (679ms) than in the incongruent condition (727ms). The factor SOA was significant with longer RTs in the 0ms SOA (762ms) than 500ms SOA (679ms; β = 97, t(20580) = 31.6, p<.0001), which in turn, had longer RTs than the 1s SOA (664ms; β = 14, t(20579) = 4.8, p<.0001). Finally, the interaction between Task and Congruency was significant (β = 30, t(20541) = 5.9, p<.0001), showing that the congruency was larger in the Same task (63ms; p<.0001) than in the Opposite task (35ms; p<.0001).

Table 4. RTs and standard deviation (SD) for each task (Same; Opposite), SOA (-1000; -500; 0) and congruency (Congruent (con.); Incongruent (Incon.)), and the congruency effect (diff) of Experiment 2.

	Same task								
	-1000			-500			0		
	Con.	Incon.	diff.	Con.	Incon.	diff.	Con.	Incon.	diff.
Bil.	623 (104)	673 (120)	50	631 (96)	719 (99)	88	722 (84)	782 (110)	60
Mon.	634 (94)	699 (97)	65	638 (81)	690 (107)	52	738(82)	800 (112)	62
	Opposite task								
	-1000			-500			0		
	Con.	Incon.	diff	Con.	Incon.	diff.	Con.	Incon.	diff.
Bil.	653 (106)	679 (113)	26	656 (92)	692 (112)	36	738 (95)	764 (107)	26
Mon.	659 (86)	708 (98)	49	692 (94)	730 (103)	38	772 (84)	796(107)	24

Note. Bil. = Bilinguals Mon. = Monolinguals

In the error analysis, systematic removal of factors from the most complex model led to one that included the factors Task, Group, and their interaction. The significant factor Task ($\beta = -0.3$, $z = 2.88$, $p = .004$) showed that more errors were made in the Opposite task (6.9%) than in the Same task (5.6%). The significant factor group ($\beta = -1.3$, $z = 3.69$, $p = .0002$) showed that overall, bilinguals made fewer errors (3.34%) than monolinguals (9.23%). The interaction between Task and Group was also significant ($\beta = 0.9$, $z = 6.76$, $p < .0001$) revealing that, Bilinguals produced fewer errors in the Opposite (2.86%) than the Same task (3.81%; $\beta = 0.3$, $z = 2.88$, $p = .02$) whereas monolinguals produced more errors in the Opposite (11.15%) than the Same task (7.44%; $\beta = -0.5$, $z = -7.52$, $p < .0001$). Further, in line with our predictions, the group difference was not significant in the Same task ($\beta = 0.4$, $z = 1.19$, $p = .6$, whereas this difference was significant in the Opposite task ($\beta = 1.3$, $z = 3.69$, $p = .001$)

6.2. Cue ERPs

N1 amplitude was significantly different between groups ($F(1,33)=6.8$, $p=.013$), showing that the N1 was more negative in amplitude in bilinguals (-1.8μV, SE=.6) than in monolinguals (.4μV, SE = .6; Fig 8). In contrast to the results of Experiment 1, the analysis on Cue N2 amplitude did not reveal any significant effects.

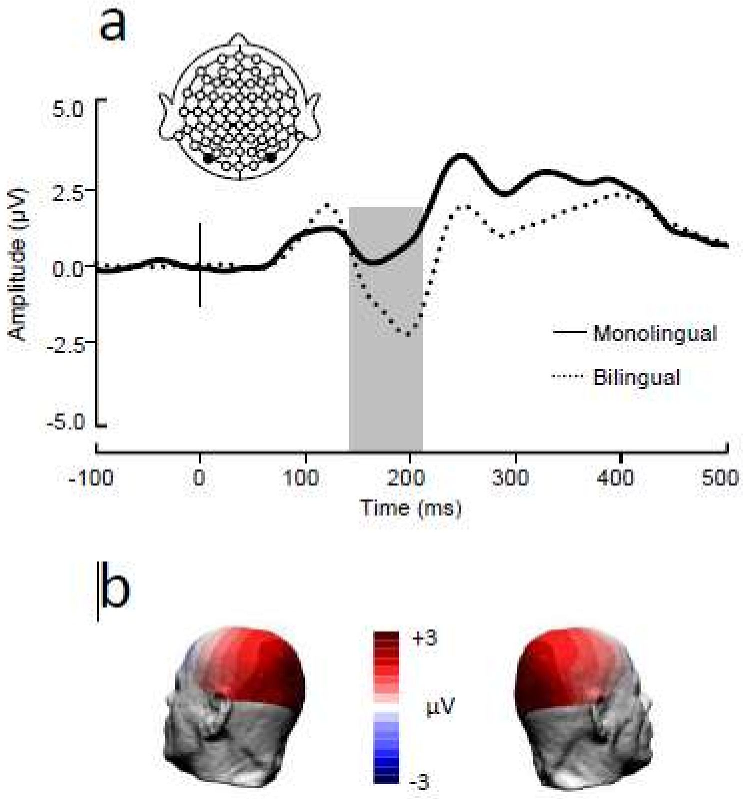

Figure 8. Grand average cue ERPs of Experiment 2

a. Grand average ERP waveforms elicited by the cue stimulus at negative SOAs. Depicted is a linear derivation of electrodes PO7 and PO8. The grey-shaded area is the time-window used to analyze the N1.

b. Topography of the group difference (monolinguals – bilinguals)

6.3. Target ERPs

The analysis in mean N1 amplitude revealed a significant main effect of SOA (F(2,66)=80, p <.001) showing that the largest N1 amplitude at 0ms (-.4µV, SE=.5) followed by 1s (-2.5µV, SE = .5), and 500ms (-4.6µV, SE = .5; all comparisons significant, all p < .001; Figure 9). The interaction between SOA and Group was also significant (F(2,66)=3.4, p = .039). Follow-up t-tests at each SOA revealed that the groups' N1 differed at the 0ms SOA only with a greater negativity for bilinguals (-1.9µV, SE=.7) than monolinguals (1.1µV, SE= .7; t(33)=2.9,p=.007).

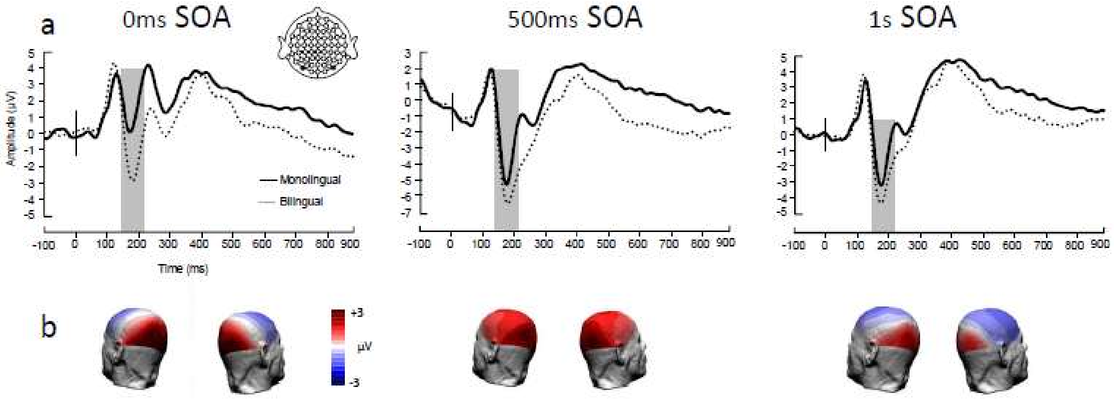

Figure 9. Grand average N1 ERP of Experiment 2

a. Grand average ERPs of electrodes PO7 and PO8 of the target stimulus of Experiment 2. The grey-shaded area is the time-window used to analyze N1 amplitude.

b. Topography of the group difference (monolinguals – bilinguals)

This pattern of the N1 was amplified in the fronto-central P2, which revealed a significant main effect of Group (F(1,33) = 5.0, p= .032) with overall larger P2 amplitude in bilinguals (5.µV, SE= .5) than monolinguals (3.4µV, SE = .5), a significant interaction between SOA and group (F(2,66) = 3.7, p= .029), as well as a significant interaction between SOA, Congruency and Group (F(2,66) = 3.2, p= .047; Figure 10). These interactions were further investigated by separate ANOVA's for each SOA. At the 0ms SOA, bilinguals displayed a larger P2 (5.3µV, SE= .5) than monolinguals (3.3µV, SE= .5; F(1,33) = 6.6, p= .014;). The 3-way interaction was likely caused by a marginally significant interaction between Congruency and Group (F(1,33) = 4.0, p= .053) at this SOA. Follow-up tests revealed that the Congruency effect was significant in monolinguals (Congruent 2.93µV, SE= .6, Incongruent 3.7µV, SE= .6; p= .032), whereas in bilinguals there was no such difference (p=.8). At the 500ms SOA, no effects with congruency or group were significant. At the 1s SOA, only the difference between groups was significant (Bilinguals 5.1 µV SE= .6, monolinguals 2.7µV SE= .7; p=.021). Therefore, the P2 was greater overall for bilinguals than monolinguals and the monolinguals' P2 distinguished stimulus congruency when they had no time to prepare for the task.

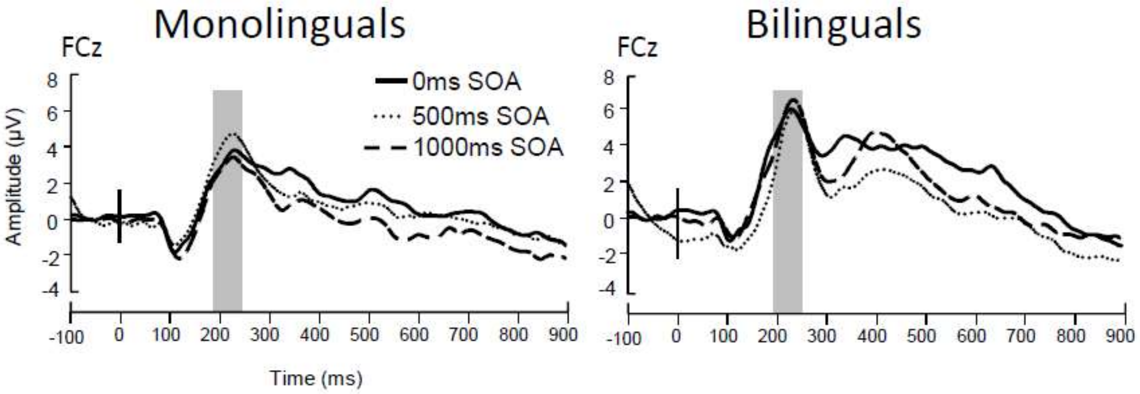

Figure 10. Grand average P2 ERP of Experiment 2

Grand average ERP waveforms of representative electrode FCz of each SOA for each group. Although P2 amplitude is larger for bilinguals than monolinguals, the P2 (highlighted in grey) at 0ms SOA is no longer suppressed as was the case in Experiment 1.

The analysis on N2 amplitude revealed a significant effect of Task (F(1,33) = 5.9, p= .021) showing that the Opposite task (2.4μV SE= .5) elicited a greater conflict monitoring response than the Same task (2.7μV SE = .5). The effect of SOA was also significant (F(2,66) = 15, p = .001) which was due to less negativity in the 0ms SOA (3.6μV SE = .5) than in both the 500ms SOA (2.2μV SE= .5; p=.003) and the 1s SOA (1.6μV SE= .5; p<.001), whereas the latter two SOAs did not differ (p=.7). Finally, analysis of the Response-locked CRN did not reveal any significant effects.

7. Discussion

The aim of Experiment 2 was to test if the pattern of results observed in Experiment 1 was due to the auditory nature of the task cue. By replacing the auditory task cue with a visual one we used an entirely visual task while keeping the task demands the same as in Experiment 1. The behavioral results were indeed not the exactly same as in experiment 1, because there was no longer a difference in RTs between groups whereas the task-difficulty manipulations had the same effect. However, bilinguals still made fewer errors overall than monolinguals and they made fewer errors in the more difficult Opposite task than the Same task. Therefore, bilinguals outperform monolinguals in these tasks and this is not solely due to an enhanced response to an auditory cue.

The ERP results are in line with this suggestion since group differences were observed in the occipital N1 and frontal P2 elicited by the cue and target stimuli. The visual N1 response to the task cue was increased in bilinguals, which given that the N1 reflects early visual categorization (Vogel & Luck, 2000), suggests bilinguals engage more neural resources in categorizing the cue. This increased negativity of the bilingual N1 was also observed for the combined target-cue stimulus at the 0ms SOA indicating the same processes are likely at play in this condition.

This pattern of the N1 was maintained at the P2, however, the monolingual P2 also distinguished between congruent and incongruent trials. This suggests monolinguals benefited from having a visual cue such that their early attention allocation distinguishes between conflicting and non-conflicting trials. Further, monolinguals no longer distinguished between tasks in their response monitoring, which suggests they are more proactive in managing task demands with the visual task cue. Finally, the conflict-monitoring N2 response to the target distinguished between tasks in both language groups which indicates that the Opposite task was perceived as the more challenging one by both language groups.

In the results of Experiment 1 and 2, we saw language group differences in behavioral and ERP measures. However, it is not clear how these measures relate to each other, and more importantly, which measures directly relate to language skills. Also, to account for the fact that some monolinguals had been exposed to another language in school settings and bilingual participants show large variability in their L1 and L2 proficiency and usage, we next correlated the language questionnaire data with the behavioral and ERP results, averaged over all experimental conditions.

8. ERP-RT-Language questionnaire correlation analyses experiments 1 & 2

In the correlation analysis we separately analyzed the average of the negative SOAs and the 0ms SOA, to evaluate the unique contributions of the cue and target, and their combined presentation. The average amplitude of all remaining conditions was calculated for each ERP component and correlated with L1 and L2 proficiency. We only report correlations between measures that span across experiment phases (cue-target-response) or types of measures (RTs-ERPs-language questionnaires) to avoid spurious correlations between e.g., highly correlated ERPs. Further, a separate analysis was performed on cue N1 amplitude of Experiment 2, because there was no visual N1 elicited by the cue in Experiment 1.

The bilinguals' language proficiency correlated negatively with RTs at the 0ms SOA (L1: R=-.46, p= .005; L2: R=-.52, p= .001), showing that greater L1 and L2 knowledge increases performance when there is no time to prepare for the tasks. Bilinguals' L2 proficiency also correlated positively with cue N2 (R=.36, p=.03) and target N2 of the 0ms SOA only (R=.36, p=.03), which shows that greater L2 proficiency is associated with a smaller conflict monitoring response. Further correlations were found between RTs and CRN (negative SOAs: (R=.40, p=.014; 0ms SOA: (R=.32, p=.054) and between target N2 and CRN (negative SOAs: (R=.-40, p=.014; 0ms SOA: (R=-.37, p=.025). This shows that in bilinguals, a greater target conflict-monitoring response and longer RTs are followed by less performance monitoring.

The only significant correlations in monolinguals were between L2 proficiency and cue N2 (R=.36, p= .037), and between Cue N2 and target P2 (R=-.44, p=.01), both in the 0ms condition only. The correlation between L2 proficiency and cue N2 amplitude was also observed in bilinguals and it therefore provides solid evidence that knowledge of another language, even when it is not much, reduces the conflict-monitoring response elicited by our task cue. Finally, the correlation between cue N2 and target P2 shows monolinguals' attention orienting response to the target is larger when preceded by a relatively large cue conflict-monitoring response.

9. General discussion

The bilingual executive control advantage has been highly contested, primarily contingent upon often-found null results in the young adult population (e.g. Paap & Sawi, 2014). Indeed, we did not find any language group differences in performance of the standard arrow flanker task. However, following the proposition that these null effects are due to the low task difficulty for a population at the peak of their cognitive capacity (Bialystok, 2016; 2017), we aimed to increase task difficulty with our Opposite flanker task and different task preparation times.

The behavioral results showed that our task modifications indeed increased task demands as evidenced by longer RTs and more errors in the conditions we predicted to be more difficult. Having established this, the next aims of the experiments were to test if bilinguals indeed outperform monolinguals in these more difficult conditions, and if so, to identify cognitive processes related to this behavior.

In line with the predictions of the task-load hypothesis, we saw bilinguals were less hampered by the more difficult task in Experiment 1 and they displayed fewer errors overall than monolinguals. However, when using a visual task cue instead of an auditory one, the RT difference between language groups was no longer present. This suggests that part of the bilingual executive control advantage in our experiments is due to superior auditory processing in bilinguals (c.f. Kritzman et al., 2014), which also explains their greater attention-orienting response and task-sensitive conflict monitoring response to auditory cues.

However, bilinguals still displayed greater perceptual ERP responses to the cue and had fewer errors than monolinguals when the task cues were visual. Therefore, superior auditory processing is not the sole source of the bilingual executive control advantage. Insight in other potential factors involved was provided by the correlation analyses. In these analyses we saw that greater L2 proficiency is associated with a smaller cue conflict-monitoring response in both language groups. This surprising finding shows that learning a second language impacts how the brain responds to conflicting stimuli, which is one way in which L2 learning induced neuroplasticity in the anterior cingulate cortex (Abutalebi et al., 2012) could be expressed. Better preparation for a task facilitates task performance and L2 proficiency appears to drive this, which was confirmed by the observation that higher L2 proficiency was also associated with shorter RTs in bilinguals.

Together, these observations provide new insight in how learning a second language influences how the brain processes stimuli. Our language groups differed qualitatively in how they deal with high task demands which supports the adaptive control hypothesis (Green & Abutalebi, 2013) in that the bilingual experience promotes proactive, flexible management of task demands in continuously varying circumstances. By contrast, monolinguals displayed later control processes in the form of response monitoring. Therefore, it appears that the bilingual executive control advantage is due to greater allocation of neural resources at early processing stages, whilst monolinguals apply more resources at later processing stages, here, even after a response was given. Such early bilingual processing differences are in line with the BAPSS model (Grundy, Anderson, & Bialystok, 2017), in which bilingual stimulus processing is assumed to rely more on early automatic processes. Importantly, the language group differences we observed become evident when task demands are high, and auditory processing, inhibition, and flexible task-set control are part of the task.

Finally, we further found that upon simultaneous presentation of audiovisual stimuli, bilinguals appeared to prioritize the auditory signal despite the visual stimulus being the target, which monolinguals appeared to prioritize. Given that the task does not rely on auditory processing and auditory prioritization was not associated with general task performance, we propose that the bilingual P2 effect is due to an automatic capture of attention of auditory stimuli. A bilingual's experience with different sets of phonological representations for each of their languages influences their auditory processing (Simonet, 2016), with more robust auditory signal encoding at the brainstem relative to monolinguals (Krizman, Marian, Shook, Skoe, & Kraus, 2012; Krizman, Skoe, Marian, & Kraus, 2014), and greater auditory attentional responses (Kuipers & Thierry, 2010; 2012; 2015). Such enhanced auditory processing characteristics are not unique to bilinguals, musicians have been found to display similar patterns (Moreno et al., 2014). However, it is unclear to what extent the amount of phonological overlap between a bilingual's languages contributes to their increased phonological processing capacity. One could argue that little overlap (e.g, between English and Chinese) may challenge phonological processing more because two entirely different phonological repertoires need to be mastered. On the other hand, more phonologically similar languages (e.g, English and Dutch) require greater attention to phonetics because shared phonemes e.g., /e/ may map onto different vowels (e.g., "i" in Dutch and "e" in English).

In a similar study to ours, Wu and Thierry (2013) found that when in a bilingual mode, their participants experienced less interference from incongruent trials in a standard arrow flanker task, expressed in fewer errors and a smaller P3 ERP in this condition. This suggests that in addition to the factors that we found are involved in greater executive control in bilinguals, bilinguals appear to be in a greater state of alertness when in a bilingual context. Usually, experiments comparing monolingual and bilingual performance are set in a monolingual context, which may therefore make bilinguals perform in a "monolingual mode", similar to the monolinguals they are compared with. Future studies may seek to investigate this further by using both monolingual and bilingual contexts in their experiments.

One of the main criticisms regarding studies of executive control in bilinguals is that positive effects are usually found in small-scale reaction time studies (less than 30 participants per language group), whereas large-scale studies (100 participants or more per group) fail to replicate these behavioral findings (Anton et al, 2019; Paap et al, 2015). Given our modest sample size of twenty participants per group, one could argue our study has an elevated high risk of type I and Type II errors (Paap, Johnson & Sawi 2017). However, the results of Experiment 1 were mostly replicated in Experiment 2, whereas the above arguments would predict such "small scale" studies would display much variance, and thus different results. Further, we recorded both behavioral measures and ERPs and used a sample size typical for ERP studies (c.f., Wu & Thierry, 2007, 2013). Finally, convergent evidence for our results comes from the correlations we found between these methods and a language history questionnaire, which reveal a language group difference in the auditory orienting response that is not reflected in group RTs, language group differences in conflict monitoring that are correlated with RTs, as well as a correlation between language proficiency measures and RTs.

Another ongoing point of discussion regarding bilingualism studies is how the language groups are defined. Different research labs typically use their own language history questionnaire which is argued to be one of the reasons some studies are not replicated. Also, some argue assessment of language proficiency by self-report alone is not recommended, because language balance may be represented incorrectly (Gollan et al. 2012). However, we believe our usage of the questionnaire was warranted since the key aspect of bilingual status is correctly measured with such questionnaires (Gollan et al. 2012).

To conclude, in the light of reported failures to replicate the bilingual executive control advantage, we show that the chance of observing behavioral differences between young adult language groups increases when task demands are high and when auditory processing is an essential component of the task. In addition, differences are more likely to be found when comparing monolinguals with highly fluent or balanced bilinguals (cf. Costa et al., 2009; Kroll & Bialystok, 2013) and the within groups' variance in L1 and L2 proficiency is taken into account, rather than only relying on a separation into two language groups. Using these methods we show that stimulus processing characteristics that give rise to greater performance by bilinguals are better auditory stimulus processing, better early stimulus identification or categorization, and more efficient conflict-monitoring.

Acknowledgements

We would like to thanks Kenneth Paap, John G. Grundy, and two anonymous reviewers for their helpful and constructive comments on a previous version of this manuscript.

References

Abutalebi, J., Brambati, S. M., Annoni, J., Moro, A., Cappa, S. F., & Perani, D. (2007). The Neural Cost of the Auditory Perception of Language Switches: An Event-Related Functional Magnetic Resonance Imaging Study in Bilinguals. *The Journal of Neuroscience, 27*, 13762 – 13769. https://doi.org/ 10.1523/jneurosci.3294-07.2007.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2012). Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring. *Cerebral Cortex*, *22*, 2076 – 2086. https://doi.org/ 10.1093/cercor/bhr287.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20,* 242 – 275. https://doi.org/ 10.1016/j.jneuroling.2006.10.003.

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism*, *19*, 689 – 698. https://doi.org/ 10.1017/S1366728916000225.

Antón E., Carreiras M., Duñabeitia J. A., (2019). The impact of bilingualism on executive functions and working memory in young adults. *PLoS ONE* 14(2): e0206770. https://doi.org/10.1371/journal.pone.0206770

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699 – 714. https://doi.org/ 10.1016/j.ecresq.2014.02.003.

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, *42*, 33 – 42. https://doi.org/ 10.1111/j.1469-8986.2005.00258.x

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48. https://doi.org/10.18637/jss.v067.i01.

Baumgart, C. Q., & Billick, S. B. (2017). Positive Cognitive Effects of Bilingualism and Multilingualism on Cerebral Function: a Review. *Psychiatric Quarterly, 89*, 1 – 11. https://doi.org/ 10.1007/s11126-017-9532-9.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the simon task. *Canadian Journal of Experimental Psychology*, *60*, 68 – 79. https://doi.org/ 10.1037/cjep2006008 .

Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism*, *5*, 517 – 534. https://doi.org/ 10.1075/lab.15040.bia.

Bialystok, E. (2017). The Bilingual Adaptation: How Minds Accommodate Experience. *Psychological Bulletin, 143*, 233 – 262. https://doi.org/ 10.1037/bul0000099.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging , and Cognitive Control : Evidence From the Simon Task. *Psychology and Aging*, *19*, 290 – 303. https://doi.org/ /10.1037/0882-7974.19.2.290.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*, 103 – 119. https://doi.org/ 10.1177/13670069050090010701.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*, 215 – 222. https://doi.org/ 10.1016/S1364-6613(00)01483-2.

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,* 221 – 235. https://doi.org/ 10.1111/nyas.13879

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 367 – 379. https://doi.org/ 10.3758/CABN.7.4.367.

Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics, 77*, 2684 – 2693. https://doi.org/ 10.3758/s13414-015-0973-7.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don’t. *Cognition*, *113*, 135 – 149. https://doi.org/ 1016/j.cognition.2009.08.001.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*, 732–744.

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*, 1752 – 1765. https://doi.org/ 10.1162/jocn_a_00817.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*, *26*, 99 – 107. https://doi.org/ 10.1177/0956797614557866.

Del Maschio, N., & Abutalebi, J. (2018). Neurobiology of bilingualism. In *Bilingual Cognition and Language* (pp. 325 – 345). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135 – 168. https://doi.org/ 10.1146/annurev-psych-113011-143750.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143 – 149. https://doi.org/ 10.3758/BF03203267.

Falkenstein, M., Hoorman, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*, 87– 107.

Fernandez, M., Acosta, J., Douglass, K., Doshi, N., & Tartar, J. L. (2014). Speaking Two Langauges Enhances an Auditory but Not a Visual Neural Marker of Cognitive Inhibition. *AIMS Neuroscience, 1*, 145 – 157. https://doi.org/ 10.3934/Neuroscience.2014.2.145.

Fernandez, M., Tartar, J. L., Padron, D., & Acosta, J. (2013). Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain and Cognition*, *83*, 330 – 336. https://doi.org/ 10.1016/j.bandc.2013.09.010.

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.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*, 1357 – 1392. https://doi.org/ 10.1152/physrev.00006.2011.

Friesen, D. C., Latman, V., Calvo, A., & Bialystok, E. (2014). Attention during visual search: The benefit of bilingualism. *International Journal of Bilingualism*, *19*, 693 – 702. https://doi.org/ 10.1177/1367006914534331.

Fuchs, E., & Flügge, G. (2014). Adult Neuroplasticity: More Than 40 Years of Research. *Neural Plasticity, e541870*, 1 – 10. https://doi.org/ 10.1155/2014/541870.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 1 – 14. https://doi.org/ 3389/fpsyg.2014.00011.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468 – 84. https://doi.org/ 10.1016/0013-4694(83)90135-9.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515 – 530. https://doi.org/ 10.1080/20445911.2013.796377.

Gollan, T. H., Weissberger, G. H. Runnqvist, E., Montoya, R. I., Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism-Language and Cognition, 15*, 594-615. https://doi.org/ 10.1017/S1366728911000332.

Grundy, J. G., & Bialystok, E. (2019). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Frontiers in Psychology, 10*, 797. https://doi.org/10.3389/fpsyg.2019.00797.

Guo, T., Liu, H., L., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *Neuroimage,56*, 300 – 2309. https://doi.org/ 10.1016/j.neuroimage.2011.03.049.

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking Brainwaves to the Brain: An ERP primer. *Developmental Neuropsychology, 27*, 183 – 215. https://doi.org/ 1207/s15326942dn2702

Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research*, *1446*, 71 – 90. https://doi.org/ 10.1016/j.brainres.2012.01.052

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences, 109*, 7877 – 7881. https://doi.org/ 10.1073/pnas.1201575109.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language, 128*, 34 – 40. https://doi.org/ 10.1016/j.bandl.2013.11.006.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition, *Journal of Cognitive Psychology, 25*, 497-514, https://doi.org/ 10.1080/20445911.2013.799170.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind. What bilinguals tell us about language processing and its consequences for cognition. *Psychology of Learning and Motivation - Advances in Research and Theory, 56,* 229 – 262. https://doi.org/ 10.10van16/B978-0-12-394393-4.00007-8.

Kuipers, J. R., & Thierry, G. (2010). Event-related brain potentials reveal the time-course of language change detection in early bilinguals. *Neuroimage, 50*, 1633 – 1638. https://doi.org/ 10.1016/j.neuroimage.2010.01.076.

Kuipers, J. R., & Thierry, G. (2012). Event-related potential correlates of language change detection in bilingual toddlers. *Developmental Cognitive Neuroscience, 2*, 97 – 102. https://doi.org/ 10.1016/j.dcn.2011.08.002.

Kuipers, J. R., & Thierry, G. (2015). Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain and Language, 149*, 27 – 32. https://doi.org/ 10.1016/j.bandl.2015.07.004.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? Psychological Bulletin, 144, 394–425. https://doi.org/10.1037/bul0000142

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*; *58*, 301–324. https://doi.org/ 10.1016/j.cortex.2014.05.001

Li, P., Sepanski, S. & Zhao, X. Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods 38***,** 202–210 (2006). https://doi.org/10.3758/BF03192770

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1000–1014.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*, 383-391.

Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia*, *66*, 157 – 169. https://doi.org/ 10.1016/j.neuropsychologia.2014.11.014.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific effects. *PLoS ONE*, *9*, 1 – 8. https://doi.org/ 10.1371/journal.pone.0094169.

Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., & Kroll, J. F. (2011).When deaf signers read English: Do written words activate their sign translations? *Cognition, 118*, 286 – 292. https://doi.org/ 10.1016/j.cognition.2010.11.006.

Morris, D. J., Steinmetzger, K., & Tøndering, J. (2016). Auditory event-related responses to diphthongs in different attention conditions. *Neuroscience Letters, 626*, 158 – 163. https://doi.org/ 10.1016/j.neulet.2016.05.002.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97 – 113. https://doi.org/ 10.1016/0028-3932(71)90067-4

Paap K. R., Anders-Jefferson R., Mason, L, Alvarado, K, & Zimiga, B. (2018). Bilingual Advantages in Inhibition or Selective Attention: More Challenges. *Frontiers in Psychology, 15.* https://doi.org/10.3389/fpsyg.2018.01409

Paap, K.R., Anders-Jefferson, R., Zimiga, B., Mason, L, & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, Simon, and spatial Stroop tasks? *Cognitive Research : Principles and Implications, 5*, https://doi.org/10.1186/s41235-020-0207-y

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232 – 258. https://doi.org/ 10.1016/j.cogpsych.2012.12.002.

Paap, K. R., Johnson, H., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265 – 278. https://doi.org/ 10.1016/j.cortex.2015.04.014

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, 1 – 15. https://doi.org/ 10.3389/fpsyg.2014.00962

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The Plastic Human Brain Cortex. *Annual Review of Neuroscience*, *28*, 377 – 401. https://doi.org/ 10.1146/annurev.neuro.27.070203.144216.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76*, 1 – 23. https://doi.org/ 10.1037/h0093840.

Picton, T. W. Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology, 37, 127–152. https://doi.org/10.1111/1469-8986.3720127.

Posner, M. I, Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information processing account of its origin and significance. *Psychological Review, 83*, 157– 171. https://doi.org/ 10.1037/0033-295X.83.2.157

Price, C. J., Green, D. W., & Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, *122*, 2221 – 2235. https://doi.org/ 1093/brain/122.12.2221.

Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: case of limited transfer. *Journal of Cognitive Psychology, 25*, 622 – 645. https://doi.org/ 10.1080/20445911.2013.821993.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sandhu, R., & Dyson, B. J. (2012). Re-evaluating visual and auditory dominance through modality switching costs and congruency analyses. *Acta Psychologica, 140*, 111– 118. https://doi.org/ 10.1016/j.actapsy.2012.04.003.

Scaltritti, M., Peressotti, F., & Miozzo, M. (2017). Bilingual advantage and language switch: What's the linkage? *Bilingualism: Language and Cognition, 20*, 80 – 97. https://doi.org/ 1017/S1366728915000565.

Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of experimental psychology, 8*1, 174 – 176. https://doi.org/ 10.1037/h0027448.

Simonet, M. (2016). The Phonetics and Phonology of Bilingualism. *Oxford Handbooks Online*, 1 – 25. https://doi.org/ 10.1093/oxfordhb/9780199935345.013.72.

Thierry, G., & Wu, Y., J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*, 12530 – 12535.

van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews*, *64*, 104 – 122. https://doi.org/ 1016/j.brainresrev.2010.03.002.

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, *18*, 2706 – 2716. https://doi.org/ 10.1093/cercor/bhn030.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity' specific to errors? *Biological Psychology, 51*, 109 – 128.

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No Evidence for Bilingual Cognitive Advantage of Four Hypotheses. *Journal of Experimental Psychology General, 145*, 2 – 46. https://doi.org/ 10.1037/xge0000120.

Vogel, E. K. & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190-203.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Wu, Y. J., Thierry, G. (2013). Fast Modulation of Executive Function by Language Context in Bilinguals. *Journal of Neuroscience, 33,* 13533-13537. https://doi.org/0.1523/JNEUROSCI.4760-12.2013

Yeung, N., Botvinick, M. M., & Cohen, J.D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931 – 959. https://doi.org/ 10.1037/0033-295X.111.4.931. 931.

Zouridakis, G., Simos, P. G., & Papanicolaou, A. C. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. *Brain Topography, 10*, 183 – 189.