

The effects of input and output modalities on language switching between Chinese and English

Wai Leung Wong¹  and Urs Maurer^{1,2} ¹Department of Psychology, The Chinese University of Hong Kong, Shatin, N. T., SAR Hong Kong, China and ²Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, N. T., SAR Hong Kong, China

Research Article

Cite this article: Wong WL, Maurer U (2021). The effects of input and output modalities on language switching between Chinese and English. *Bilingualism: Language and Cognition* 24, 719–729. <https://doi.org/10.1017/S136672892100002X>

Received: 15 April 2019
Revised: 21 January 2021
Accepted: 21 January 2021
First published online: 17 March 2021

Keywords:

language control; input modality; output modality; language switching

Address for correspondence:

Urs Maurer
Associate Professor
Rm 325, Sino Building
The Chinese University of Hong Kong
Email: umaurer@psy.cuhk.edu.hk

Abstract

Language control is important for bilinguals to produce words in the right language. While most previous studies investigated language control using visual stimuli with vocal responses, language control regarding auditory stimuli and manual responses was rarely examined. In the present study, an alternating language switching paradigm was used to investigate language control mechanism under two input modalities (visual and auditory) and two output modalities (manual and vocal) by measuring switch costs in both error percentage and reaction time (RT) in forty-eight Cantonese–English early bilinguals. Results showed that higher switch costs in RT were found with auditory stimuli than visual stimuli, possibly due to shorter preparation time with auditory stimuli. In addition, switch costs in RT and error percentage could be obtained not only in speaking, but also in handwriting. Therefore, language control mechanisms, such as inhibition of the non-target language, may be shared between speaking and handwriting.

Introduction

It is well-documented that language control is common in word production of bilinguals (Green, 1998). As words from two languages are stored in bilinguals' lexicon, it is important to inhibit bilingual language production within the target language to avoid the production of the non-target language, and this inhibition is called language control (Declerck & Philipp, 2015). Language switching was considered as a way to investigate the mechanism of language control in many previous research studies (Declerck, Stephan, Koch & Philipp, 2015b; Gollan & Ferreira, 2009; Meuter & Allport, 1999; Prior & Gollan, 2011; Thomas & Allport, 2000), and the performance of language switching was measured by switch cost, a marker for language control (Declerck et al., 2015b). In a typical language switching paradigm, there are repetition trials, in which participants report a stimulus in the same language as the previous trial, and switch trials, in which participants report a stimulus in a different language from the previous trial. Switch cost is indicated by longer reaction time (RT) or higher error percentage in switch trials than repetition trials. In other words, switch cost is calculated by subtracting the RT or error percentage on repetition trials from those on switch trials (Declerck et al., 2015b).

One of the influential models that explains the mechanism of language control is the inhibitory control model (ICM; Green, 1998). According to the ICM, when bilinguals process a concept, parallel activation of both relevant languages is induced, which is followed by the inhibition of the non-target language. During the repetition trials, the non-target language is kept inhibited. However, during the switch trials, the previously inhibited language has to be produced, and therefore the inhibition of the non-target language has to be overcome. Therefore, longer processing time is required and it signifies the switch cost. This model posits that language control is a process that is situated at the lexical-semantic level.

Despite extensive research on language control, this topic was mainly investigated by studies adopting language switching paradigms with visual stimuli (image or words) and vocal production (speaking) (e.g., Gollan & Ferreira, 2009; Meuter & Allport, 1999; Prior & Gollan, 2011; Thomas & Allport, 2000), while auditory input (sound) and manual output (handwriting) modalities were neglected. Therefore, while language switching has been used to investigate language control for several decades, little is known about whether the inhibition mechanism can be generalized to listening and writing. However, it is theoretically interesting to investigate the effects of different modalities on language control because, in daily conversation, we receive auditory input and generate vocal output (Kaufmann, Mittelberg, Koch & Philipp, 2018). In addition, language switching in writing is common, although less frequent than in daily conversation (Yau, 1993). Additionally, the investigation of different modalities and language switching can enhance our understanding of the language control mechanism by exploring whether the inhibition mechanism used in language switching in speaking can be

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE
UNIVERSITY PRESS

applied to handwriting. Hence, language control under different modalities was investigated in the present study by using two input modalities (visual and auditory stimuli) and output modalities (manual and vocal word production).

Although most research studies related to language switching adopted visual stimuli, there was a recent study comparing the switch costs between visual and auditory input modalities, with vocal word production as output modality (Declerck et al., 2015b). In this study, the average RT of vocal production was found to be longer after receiving auditory inputs (sound) than visual inputs (images), but a reversed pattern was shown for switch costs, which were higher in the visual condition than the auditory one. As a previous study has shown that semantic priming was stronger and lasted longer with auditory stimuli than visual stimuli (Holcomb & Neville, 1990), Declerck et al. (2015b) explained the longer RT in the auditory condition than the visual condition by longer lexical-semantic processing with auditory stimuli than visual stimuli. The longer RT with auditory stimuli resulted in longer inter-stimulus interval (ISI) and this gave rise to the potential decay effect, which is the dissipation of activation of the previous language representation over time. Therefore, lower switch costs were shown in the auditory condition (Declerck et al., 2015b). However, as this is the only study which compared language switch costs between two input modalities, more studies are needed to confirm this proposition.

Although most of the research studies regarding language processing were about word production in speaking instead of handwriting, the process of word production in handwriting was thought to be similar to that in speaking. A model proposed by Bonin, Roux, Barry and Canell (2012) illustrated that after a person perceives an image or a sound, he or she will conduct perceptual analysis for stimulus recognition. Then, the person will enter the stage of conceptual and semantic processing, which is shared by both speaking and handwriting. However, the process then diverges into phonological L-level when speaking, and orthographic L-level when handwriting during the stage of word-form encoding.¹ Finally, a phoneme level for speaking and a grapheme level for writing are achieved respectively, and words can be produced vocally or manually. The similarity and difference of word production in handwriting and speaking were further supported by an EEG study by Perret and Laganaro (2012), which has shown that the conceptual and lexical-semantic processes between speaking and handwriting were shared, but phonological and orthographic word-form encoding processes were different.

Despite the word-form encoding difference between speaking and writing, some previous studies have investigated language control in signing (Kaufmann et al., 2018; Schaeffner, Fibla & Philipp, 2017) and typing (Schaeffner et al., 2017), and they have shown that the shared phonological information between writing and speaking might play a main role in language control. By comparing unimodal (spoken-spoken) and bimodal (experiment 1: spoken-signed; experiment 2: spoken-typed) switching, Schaeffner et al. (2017) found that switch costs were lower in bimodal switching than unimodal switching only when participants had to switch between speaking and signing, but not between speaking and typing. The output channels of speaking and signing or typing were the mouth and hands respectively. During bimodal switching, although lemmas in both languages can remain uninhibited at the lexical level as people can produce both languages simultaneously, the irrelevant output channel may need to be inhibited to ensure that a correct output channel is being used. The study by Schaeffner et al. (2017) demonstrated

that language inhibition in switching between speaking and typing was more costly than output channel inhibition in switching between speaking and signing. This was explained by the phonological overlap between speaking and typing: as typing and handwriting share similar retrieval mechanism of phonological information (Pinet, Ziegler & Alario, 2016; Schaeffner et al., 2017), but there is no phonological overlap between speaking and signing, for phonological information may not be required in signing. According to the phonological mediation hypothesis, writing consists of inner speech (e.g., Geschwind, 1969; Luria, 1970), and it demands the same phonological information as speaking (Schaeffner et al., 2017). This was supported by a previous study showing that inconsistent spelling will induce more writing errors, compared with consistent spelling (Bonin, Peereman & Fayol, 2001). Therefore, phonological information may play an important role in both writing and speaking. Due to this reason, we hypothesized that switch cost can be found in language switching in handwriting.

While previous studies about modalities and language switching were scarce, the relationship between modalities and task switching was investigated more extensively. For example, a classic task-switching study has manifested that adopting visual input and manual output in the same task overloaded the visuospatial sketch pad, a mechanism specialized for the short-term storage of spatial information (Logie, Zucco & Baddeley, 1990), thereby creating interference (Brooks, 1968). However, another study has shown that using compatible modalities (i.e., auditory-vocal and visual-manual) created less interference than incompatible modalities (i.e., auditory-manual and visual-vocal), due to our natural tendencies to bind certain stimuli to certain responses (Stephan & Koch, 2010), and the cross-talk at the level of modality-specific processing pathways (Stephan & Koch, 2011). Input-output modality compatibility was defined as the similarity of stimulus modality and modality of response-related sensory consequences (Stephan & Koch, 2011). Due to the different findings discovered by previous studies, the interaction of modalities on task switching has yet to be confirmed.

Although aforementioned studies were about task switching, the inhibition mechanism of task switching may be similar to that of language switching (Meuter & Allport, 1999). For example, in the Stroop task, word naming needs to be inhibited to produce the colour of the word. Similarly, L1 (or L2) needs to be inhibited to produce L2 (or L1) in language switching (Meuter & Allport, 1999). However, a recent fMRI study has shown that while task switching and language switching might share some aspects of executive control, bilinguals had higher efficiency on maintaining inhibition of a non-target language than a non-target task (Weissberger, Gollan, Bondi, Clark & Wierenga, 2015). This might be due to their frequent inhibition of language in single language contexts, but less experience on task switching. Due to the different aforementioned findings, whether there is an interaction between modalities and language switching pattern, as similar to the interactions between modalities and task switching pattern, is still unknown.

To investigate whether the results of language switching in previous studies can be generalized to auditory and manual modalities, four conditions with different input and output modalities were included in the present experiment – namely, auditory-vocal, auditory-manual, visual-manual and visual-vocal conditions. Similar to the research design of Declerck et al. (2015b), images and sound were presented in the visual and auditory conditions respectively, and speaking was required in the vocal conditions.

However, we adopted a novel way to investigate language control processing in the manual conditions by asking participants to write on a paper placed on a tablet.

An alternating, predictable language switching sequence without any cue (i.e., L1-L1-L2-L2-L1-L1 or L2-L2-L1-L1-L2-L2) was used in the present study. While some previous studies have used an unpredictable language switching sequence with a cue to indicate the required language in each trial (Meuter & Allport, 1999; Prior & Gollan, 2011), the advantage of using a predictable language switching sequence is that it can avoid any distraction from visual or auditory cues which signal the required language as it may affect the RT. However, the predictability of languages may be a confounding variable as it may affect switch costs. Nonetheless, a previous study has shown that switch costs did not change depending on the predictability of the language sequence as long as the concepts were not predictable (Declerck, Koch & Philipp, 2015a). Hence, as only languages but not concepts were predictable in the present study, switch cost should reflect language control solely.

To the best of our knowledge, there has not been any research study comparing seeing, listening, speaking and handwriting in a unimodal language switching context yet. It is important to note that in this study, we would investigate the language switching in a unimodal switching context only. In other words, participants either look at a picture or listen to some sound, and perform spoken or written word production in each trial, and participants are told about the required modalities before each block. Therefore, no modality cue is needed. In contrast, in a bimodal switching context, both inputs can be shown on the screen and both languages can be produced simultaneously, so modality cues may be required. The reason for using a unimodal design is that our primary interest is the effect of modalities on switch costs in language control, while the difference of bimodal and unimodal switching was investigated in previous studies (see Kaufmann et al., 2018; Schaeffner et al., 2017). Investigating different modalities (visual, auditory, manual and vocal) in a unimodal switching context is important because it may not only clarify the effect of auditory inputs on language control and extend the inhibition mechanism of language control into a new manual modality (writing), but also provide insight into whether different modalities affect language control mechanism in the same participants.

Based on the aforementioned studies, the present study aims to investigate the language control mechanism in different modalities by recruiting Hong Kong bilinguals, whose L1 is Chinese, a morphosyllabic language, and L2 is English. Specifically, three hypotheses were proposed. First, switch costs between input modalities (visual and auditory) were examined. Higher switch costs with visual stimuli than auditory stimuli were predicted, based on the study by Declerck et al. (2015b). Furthermore, switch costs between output modalities (vocal and manual) were compared. As speaking and writing share the same phonological information retrieval mechanism, we predicted there will be no switch cost difference between writing and speaking. Additionally, the interactions between the four modalities were explored to clarify the effects of modality compatibility on language switch costs. Specifically, we were curious about whether compatible modality creates facilitation or inhibition on switch costs, compared with incompatible modality. Based on previous studies (Declerck et al., 2015b; Stephan & Koch, 2010), we predicted that modality compatible tasks (i.e., auditory-vocal and visual-manual) should induce lower switch costs than modality incompatible tasks (i.e., auditory-manual and visual-vocal).

Table 1. Mean and standard deviation (in parenthesis) of participants' language background (N = 48).

Language Background	Chinese (L1)	English (L2)
Age of acquisition (AoA)***	0 (0)	2.81 (1.57)
Self-rated speaking proficiency*** (1: very bad, 7: very good)	6.25 (.84)	4.88 (.91)
Self-rated writing proficiency (1: very bad, 7: very good)	5.25 (1.12)	4.98 (.91)
Self-rated listening proficiency*** (1: very bad, 7: very good)	5.88 (1)	5.04 (.99)
Self-rated reading proficiency** (1: very bad, 7: very good)	5.63 (1.02)	5.08 (.9)
Self-rated overall proficiency*** (1: very bad, 7: very good)	6.73 (.57)	5.17 (.86)
Years of formal education	16.44 (1.98)	16.29 (1.83)
Percentage of daily use currently***	71.29 (14.17)	20.06 (9.96)
Years of use***	18.79 (1.91)	17.06 (2.25)
Language switching frequency in daily conversation (1: low frequency, 7: high frequency)	4.83 (1.37)	
Language switching frequency in writing (1: low frequency, 7: high frequency)	3.67 (1.39)	
Language switching frequency in texting (1: low frequency, 7: high frequency)	5.25 (1.41)	

Note. Paired sample t-test was conducted for the difference of each variable under language background between Chinese and English except for the frequency of language switching between Chinese and English, * $p < .05$, ** $p < .01$, *** $p < .001$.

Method

Participants

Forty-eight undergraduate students (19 males and 29 females) studying at the Chinese University of Hong Kong (CUHK) participated in the experiment.ⁱⁱ Their age range was 18 to 24 years ($M = 19.75$ years). According to the questionnaire administered after the experiment, all participants were native Cantonese speakers with English as their L2. Their language background examined by the questionnaire (see Procedure below) is summarized in Table 1. All participants were early bilinguals and proficient in both Chinese and English. They had received formal education in Chinese and English and have used both languages for more than twelve years. All participants received credit points of a fundamental psychology course for their participation.

Apparatus and stimuli

Ten concepts were used in this study that corresponded to one-syllabic words in both Chinese and English (See Appendix). The words were frequent in both languages (character frequency in Chinese: 301 per million; word frequency in English: 77 per million). During the experiment, the ten concepts were expressed in the form of visual stimuli (pictures) and auditory stimuli (sounds) respectively. For example, for the concept "dog", participants saw a picture depicting a dog in the visual conditions (including visual-manual and visual-vocal conditions), and heard a bark in the auditory conditions (including auditory-

manual and auditory-vocal conditions). The visual stimuli were pictures adopted from Snodgrass and Vanderwart (1980), while the duration and intensity of auditory stimuli were controlled to three seconds and 70 decibels respectively by using Praat, a computer software used for speech analysis (Boersma & Weenink, 2019). The auditory stimuli were presented in a comfortable volume during the experiment.

In all conditions, E-prime 3 was used for stimulus presentation. In the visual conditions, no extra apparatus was needed besides the computer and the computer screen. In the auditory conditions, a headphone was used for sound presentation and the computer screen was switched off to avoid visual distraction. In the vocal conditions, a Serial Response box (SR box) was connected to E-Prime for the collection of RT, which was defined as the duration between the onset of stimulus presentation and the commencement of vocal responses. In addition, a microphone was connected to the SR box for voice detection. Accuracy was recorded on-site by the experimenter. In the manual conditions, a graphic tablet (WACOM Intuos Pro Large PTH-851, with an Intuos inking contact pen) was connected to the computer for the writing detection. A white sheet of paper was placed on the tablet to enhance ecological validity because it is more common to write on a paper than on a tablet. Each paper contained forty black lines for writing. The writing RT, which was defined as the duration between the onset of stimulus presentation and the start of manual responses (the moment that the pen was first in contact with the paper, or the onset of the first stroke), was recorded by E-prime, while the accuracy was recorded later by the experimenter, based on the words written on the paper.

Procedure

At the beginning, there was a five-minute familiarization phase about the association between the stimuli and concepts used in the experiment. In this phase, participants looked at the image and listened to the sound of each concept simultaneously, with the corresponding traditional Chinese character and English word shown below the image. No time limit was set so participants could look at and listen to each stimulus as long as they wished, until they pressed any key on the keyboard to move on to the next concept. In addition, they could choose to look at and listen to the stimuli again until they were familiar with the ten concepts. Although the representation of the stimuli was linked closely to the concept based on the results of pilot tests, stimuli needed to be familiarized to ensure that participants named the concept at the basic level but not at superordinate or subordinate levels. For example, the participant was expected to write or say “drum” (basic level) when listening to the drum sound or looking at a picture of a drum, but not “instrument” (superordinate level) or “bass drum” (subordinate level).

The familiarization phase was followed by a behavioural task, in which the four conditions – namely, visual-manual, visual-vocal, auditory-manual, and auditory-vocal – were presented in separate blocks. The order of the four conditions was counterbalanced across participants. Prior to each condition, both visual and verbal instructions were given, with the emphasis on speed and accuracy. They were advised to take a break between each block, as no break was allowed within a block. Each block started with ten practice trials, which covered all concepts, followed by 40 trials in the main task. A pseudo-randomized list was created for each condition, and within each list, each concept appeared twice in Chinese, and twice in English, and each concept appeared equally

often in switch and repetition trials. In each block, participants were required to switch between Chinese and English in a predictable sequence (i.e., L1-L1-L2-L2-L1-L1 or L2-L2-L1-L1-L2-L2, counterbalanced between participants). The use of two sequences ensured that the total amount of repetition trials and switch trials were the same in each language (Declerck et al., 2015b).

In the vocal conditions, participants were asked to report the concept verbally towards a microphone as accurately and as quickly as possible when they were listening to an auditory stimulus (auditory-vocal condition) or looking at a visual stimulus (visual-vocal condition). In each trial of both conditions, participants first heard a beep sound or saw a fixation cross (+) at the centre of the screen for one second respectively, followed by the auditory or visual stimulus. The duration of stimuli was fixed at three seconds, so they did not disappear even when the participants made their responses within the stimulus duration. The stimulus was followed by a break (silence or white screen) for one second, followed by the next trial.

In the manual conditions, participants were asked to write down the concept on a white paper placed on the tablet as accurately and as quickly as possible when they were listening to an auditory stimulus (auditory-manual condition) or looking at a visual stimulus (visual-manual condition). In each trial of the manual conditions, the sequence and durations of the fixation cross or beep sound, stimulus presentation and break were the same as the vocal conditions mentioned above. Participants were required to write the words in a predictable sequence, as mentioned above (i.e., L1-L1-L2-L2-L1-L1 or L2-L2-L1-L1-L2-L2, counterbalanced between participants).

After the behavioural task, a questionnaire concerning the language background of participants was administered, including variables such as age of acquisition (AoA), years of formal education, self-rated language proficiency, years of use, and percentage of daily use of both Chinese and English. It is well-documented that self-rated score is a good indication of second language proficiency (Leblanc & Painchaud, 1985). The questionnaire also included questions related to language switching frequency between Chinese and English in daily conversation, writing and texting, with foreign language background and demographic information. Participants were then debriefed and dismissed. The whole experiment lasted about 45 minutes.

Data analyses

Data from first trials in each condition (2.5%) were discarded as they did not belong to either repetition trial or switch trial. Trials with technical errors (3.09%) were also excluded from analysis. Regarding error percentage analysis, trials with lexical errors (1.25%) and language switching errors (1.2%), including those with self-correction and hesitation, were taken into account. For the RT analysis, the errors mentioned above, and trials following an error without self-correction (3.1%) were excluded. After that, RTs with two standard deviations above or below the mean in each condition were considered as outliers and they were discarded (5.26%). Regarding the RT analysis, the remaining repetition observations (sample size \times trial number) of visual, auditory, manual and vocal modalities were 1734, 1602, 1725 and 1611, while the remaining switch observations of the four conditions were 1625, 1463, 1575 and 1513 respectively. No participant was excluded due to removal of trials.

The independent variables were all within-subject factors, including input modality (auditory vs. visual), output modality

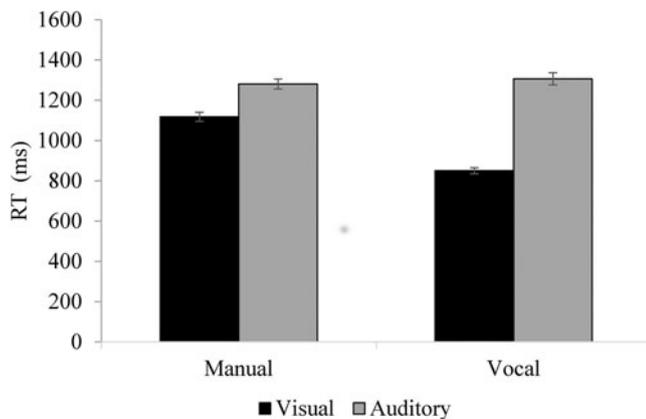


Fig. 1. Average RTs in ms in different modalities (error bars represent standard errors).

(manual vs. vocal), language (Cantonese as L1 vs. English as L2) and language transition (repetition trial vs. switch trial). The dependent variables included error percentage (for measuring accuracy) and RT. Planned comparisons t-tests will be conducted to investigate whether switch costs in RT exist in handwriting as this is the main interest of the present study.

Results

Switch costs in RT

A repeated measures analysis of variance (ANOVA) was conducted for RT, with four within-subject factors, including INPUT MODALITY (visual vs. auditory), OUTPUT MODALITY (manual vs. vocal), LANGUAGE (L1 vs. L2) and LANGUAGE TRANSITION (repetition vs. switch). The RTs in the auditory conditions (1294 ms, 95% confidence interval (CI) = 1249, 1339) were significantly longer than those in the visual conditions (984 ms, CI = 950, 1017; *input modality*, $F(1, 47) = 271.5$, $p < .001$, $\eta_p^2 = .852$), which was more pronounced in the vocal than the manual conditions (*input modality* \times *output modality*, $F(1, 47) = 85.33$, $p < .001$, $\eta_p^2 = .645$; see Figure 1). Moreover, longer RTs were found in the manual (1199 ms, CI = 1159, 1239) than the vocal conditions (1079 ms, CI = 1036, 1121; *output modality*, $F(1, 47) = 32.06$, $p < .001$, $\eta_p^2 = .406$). In addition, RTs were significantly longer in the L1 trials (1156 ms, CI = 1120, 1191) than the L2 trials (1122 ms, CI = 1085, 1158; *language*, $F(1, 47) = 15.25$, $p = .001$, $\eta_p^2 = .245$), which was more obvious in the manual conditions than the vocal ones (*output modality* \times *language*, $F(1, 47) = 4.4$, $p < .01$, $\eta_p^2 = .086$), as shown in Figure 2. Importantly, RTs of the switch trials (1165 ms, CI = 1129, 1201) were significantly longer than that of the repetition trials (1112 ms, CI = 1076, 1149; *language transition*, $F(1, 47) = 34.84$, $p < .001$, $\eta_p^2 = .436$), which was more pronounced in the auditory trials than that in the visual trials (*input modality* \times *language transition*, $F(1, 47) = 6.71$, $p < .05$, $\eta_p^2 = .125$), as shown in Figure 3. In other words, higher switch costs in terms of RT were found with auditory than visual stimuli. No other significant interactions were found. Average switch costs, with the mean RTs of switch trials and repetition trials, in different modalities are shown in Table 2, and the left part of Table 3 shows the full statistics of the above four-way repeated measures ANOVA.

Post-hoc paired-sample t-tests were conducted to clarify the three two-way interactions we found in the ANOVA. Regarding

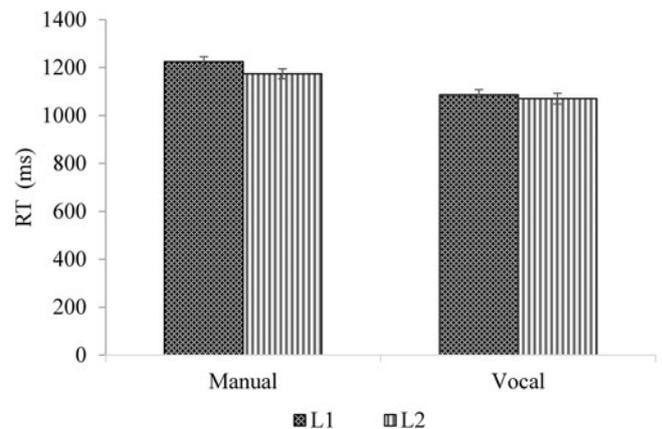


Fig. 2. Interaction of RTs between output modalities and language (error bars represent standard errors).

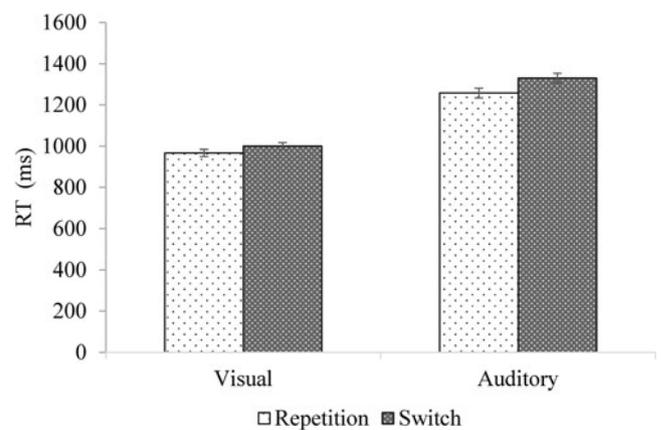


Fig. 3. Interaction of RTs between input modalities and language transition (error bars represent standard errors).

the interaction of *input modality* and *output modality*, the t-test results showed that longer RTs with auditory stimuli than visual stimuli were found in both vocal conditions (454 ms, CI = 397, 511; $t(47) = 16.14$, $p < .001$, $d = 2.72$) and manual conditions (175 ms, CI = 125, 223; $t(47) = 7.15$, $p < .001$, $d = 1$). However, for the interaction of *output modality* and *language*, the t-test results showed that the significantly longer RTs in L1 trials than L2 trials were found in the manual conditions only (51 ms, CI = 24, 77; $t(47) = 3.85$, $p < .001$, $d = .35$), but not in the vocal conditions (17 ms, CI = -4, 37; $t(47) = 1.64$, $p = .11$, $d = .11$). In addition, t-test results related to the interaction of *input modality* and *language transition* showed that switch costs were found in both auditory conditions (74 ms, CI = 42, 107; $t(47) = 4.55$, $p < .001$, $d = .43$) and visual conditions (34 ms, CI = 21, 48; $t(47) = 5.07$, $p < .001$, $d = .31$).

A planned comparisons paired-sample t-test was conducted to investigate whether switch cost exists in handwriting by comparing the RTs in switch trials and repetition trials. Significant results with longer RTs in switch trials than repetition trials were found in both visual-manual ($t(47) = 5.016$, $p < .001$, $d = .284$, CI = 28, 66) and auditory-manual conditions ($t(47) = 3.315$, $p < .01$, $d = .404$, CI = 33, 134). It showed that switch cost existed in handwriting, notwithstanding input modalities.

As there were large RT differences between different modalities, we calculated the proportional switch cost by dividing the

Table 2. Mean and standard deviation (in parenthesis) of RT in ms in switch and repetition trials and the ensuing switch costs in different input modalities and output modalities.

Input modality	Output modality					
	Manual			Vocal		
	Switch	Repetition	Switch cost	Switch	Repetition	Switch cost
Visual	1131 (164)	1084 (168)	47 (65)	865 (113)	844 (125)	21 (72)
Auditory	1324 (228)	1240 (181)	83 (174)	1342 (221)	1276 (217)	65 (140)

Table 3. Statistical results of four-way repeated measures ANOVA (input modality x output modality x language x language transition) of RT and error percentage.

Factors	RT			Error Percentage		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Input modality	271.5	< .001	.852	7.89	.007	.144
Output modality	32.06	< .001	.406	< 1	.771	.002
Language	15.25	< .001	.245	< 1	> .99	< .001
Language transition	34.84	< .001	.426	13.08	.001	.218
Input modality x output modality	85.32	< .001	.645	3.27	.077	.065
Input modality x language	2.23	.142	.045	< 1	.507	.009
Output modality x language	4.4	.041	.086	1.26	.267	.026
Input modality x output modality x language	< 1	.814	.001	< 1	.689	.003
Input modality x language transition	6.71	.013	.125	2.61	.113	.053
Output modality x language transition	1.11	.298	.023	< 1	.699	.003
Input modality x output modality x language transition	.43	.513	.009	1.699	.199	.035
Language x language transition	1.53	.222	.032	< 1	.574	.007
Input modality x language x language transition	< 1	.494	.01	6.99	.011	.129
Output modality x language x language transition	< 1	.876	.001	6	.018	.113
Input modality x output modality x language x language transition	< 1	.884	< .001	.247	.622	.005

Note. The significant results ($p < .05$) are in bold.

average switch costs in RT by average RT of repetition trials in different conditions (Declerck et al., 2015b), and a repeated measures ANOVA was conducted for the proportional switch costs with three within-subject factors, including *input modality* (visual vs. auditory), *output modality* (manual vs. vocal) and *language* (L1 vs. L2). The only significant result was found in *input modality*, with higher switch costs with auditory stimuli (7.1%, CI = 4.6, 9.6) than that with visual stimuli (4.1%, CI = 2.5, 5.8; *input modality*, $F(1, 47) = 5.19$, $p < .05$, $\eta_p^2 = .099$). No other main effect (*output modality*, $F(1, 47) < 1$, $p = .464$, $\eta_p^2 = .011$; *language*, $F(1, 47) = 1.11$, $p = .299$, $\eta_p^2 = .023$) nor interaction (*input modality x output modality*, $F(1, 47) = .106$, $p = .746$, $\eta_p^2 = .002$; *input modality x language*, $F(1, 47) = .31$, $p = .58$, $\eta_p^2 = .007$; *output modality x language*, $F(1, 47) = .051$, $p = .823$, $\eta_p^2 = .001$; *input modality x output modality x language*, $F(1, 47) = .013$, $p = .91$, $\eta_p^2 < .001$) were found.

Switch costs in error percentage

Similarly, a repeated measures ANOVA was conducted for error percentage, with the same within-subject factors – namely, *input modality* (visual vs. auditory), *output modality* (manual

vs. vocal), *language* (L1 vs. L2) and *language transition* (repetition vs. switch). Participants made more errors when the stimuli were presented auditorily (3.7%, CI = 2.6, 4.8) than visually (1.7%, CI = .7, 2.6; *input modality*, $F(1, 47) = 7.89$, $p < .007$, $\eta_p^2 = .144$). In contrast, no significant difference was found between error percentages in the vocal conditions (2.6%, CI = 1.6, 3.6) and the manual conditions (2.8%, CI = 1.8, 3.8; *output modality*, $F(1, 47) < 1$, $p = .771$, $\eta_p^2 = .002$). In addition, no significant difference was found between the error percentage in L1 (2.7%, CI = 1.8, 3.5) and L2 (2.7%, CI = 1.8, 3.5; *language*, $F(1, 47) < 1$, $p = .995$, $\eta_p^2 < .001$). However, participants made more errors in the switch trials (3.4%, CI = 2.5, 4.2) than the repetition trials (2%, CI = 1.2, 2.8; *language transition*, $F(1, 47) = 13.08$, $p = .001$, $\eta_p^2 = .218$), indicating the existence of switch costs reflected by error percentage. Moreover, an interaction was found between *input modality*, *language* and *language transition* ($F(1, 47) = 6.99$, $p < .05$, $\eta_p^2 = .129$). In addition, there was an interaction between *output modality*, *language* and *language transition* ($F(1, 47) = 6$, $p < .05$, $\eta_p^2 = .113$). The right half of Table 3 shows the full statistics of the above four-way repeated measures ANOVA. Average switch costs in error percentage in different modalities are shown in Table 4.

Table 4. Mean and standard deviation (in parenthesis) of error percentage in different input modalities, output modalities and language transitions.

Input modality	Output modality					
	Manual			Vocal		
	Switch	Repetition	Switch cost	Switch	Repetition	Switch cost
Visual	1.19% (2.13%)	1.15% (3.61%)	.04% (2.77%)	2.73% (6.57%)	1.5% (6.2%)	1.23% (3.71%)
Auditory	5.67% (7.63%)	3.13% (5.75%)	2.54% (7.32%)	3.98% (5.27%)	2.13% (4.15%)	1.85% (6.22%)

Post-hoc paired-sample t-tests of error percentage switch costs in different languages and input modalities were conducted to clarify the interaction between *input modality*, *language* and *language transition*. The only significant difference was found between visual L2 error switch cost (.06%) and auditory L2 error switch cost (3.14%; $t(47) = 3.016$, $p < .01$, $CI = 1.02, 5.1$). No other significant difference was found (visual L1 vs. L2 error switch cost: $t(47) = 1.727$, $p = .091$; auditory L1 vs. L2 error switch cost: $t(47) = 1.682$, $p = .099$; visual vs. auditory L1 error switch cost: $t(47) = .221$, $p = .826$; visual L1 vs. auditory L2 error switch cost: $t(47) = 1.638$, $p = .108$; visual L2 vs. auditory L1 error switch cost: $t(47) = .78$, $p = .439$).

Additionally, post-hoc paired-sample t-tests of error percentage switch costs in different languages and output modalities were conducted to clarify the interaction between *output modality*, *language* and *language transition*. The only significant difference was found between manual L1 error switch cost (.34%) and manual L2 error switch cost (2.13%; $t(47) = 2.1$, $p < .05$, $CI = .08, 3.5$). No other significant difference was found (vocal L1 vs. L2 error switch cost: $t(47) = .816$, $p = .419$; manual vs. vocal L1 error switch cost: $t(47) = 1.829$, $p = .074$; manual vs. vocal L2 error switch cost: $t(47) = 1.16$, $p = .252$; manual L1 vs. vocal L2 error switch cost: $t(47) = .843$, $p = .404$; manual L2 vs. vocal L1 error switch cost: $t(47) = .143$, $p = .887$).

To summarize, the ANOVA results regarding RT showed that overall switch costs were found, and they were higher with auditory than visual stimuli. The planned comparisons t-test showed that switch costs occurred in writing with both input modalities. In addition, switch costs also were found in terms of error percentage. However, the interpretation of the two three-way interactions in the error percentage analysis may need to be cautious due to the power issue (see discussion below). According to the post-hoc t-tests, RTs were longer with auditory than visual stimuli, which was more obvious in speaking than writing. Moreover, RTs were longer in L1 than L2 trials in writing only, but not speaking.

Discussion

The aim of the present study was to investigate language control by measuring switch costs in different input and output modalities. To this end, we tested bilingual Chinese–English speakers in a picture and sound naming experiment in which two input (auditory and visual) and two output modalities (vocal and manual) were combined in four separate blocks. As expected, there was no difference on switch costs between speaking and writing. However, unexpectedly, higher switch costs were found with auditory than visual stimuli, and there was no switch cost difference between compatible and incompatible modalities. The results will be discussed in detail below.

The result that switch costs could be found in auditory and manual modalities has theoretical implications for language control in bilingualism. It suggests that language control mechanisms, such as inhibition, may be used for language switching in general, no matter whether participants receive visual or auditory information, or produce spoken or written words. This suggests for example that Green's inhibitory control model (ICM; Green, 1998) can be generalized to the modality of handwriting. Accordingly, our findings suggest that this inhibition mechanism does not only apply in visual and vocal modalities, but also in auditory and manual modalities. Similar to vocal word production, language control in manual word production may also require the inhibition of non-target language at the lexical-semantic level so that bilinguals are able to switch their languages successfully. This can be explained by shared lexical-semantic and phonological retrieval processes between speaking and handwriting. It is in line with the results by a previous study that more language inhibition was required in switching between speaking and typing, compared with speaking and signing (Schaeffner et al., 2017), as the phonological retrieval processes were more overlapping between speaking and typing, compared with speaking and signing.

Support for an inhibition mechanism in language control comes from an L1 global slowing effect in our handwriting data. It has been suggested that longer RTs in L1 trials than L2 trials reflect more sustained inhibition for L1 (Bobb & Wodniecka, 2013). Although this measure did not provide unequivocal evidence that inhibition is the only mechanism under language control, it shows the possibility that inhibition may explain at least part of the language control processing (see Declerck & Philipp, 2015 for discussion about other two measures of inhibition mechanism: switch cost asymmetry and n-2 language repetition costs). This L1 global slowing effect was also found in speaking in some previous studies (e.g., Christoffels, Firk & Schiller, 2007; Costa & Santesteban, 2004; Costa, Santesteban & Ivanova, 2006; Gollan & Ferreira, 2009; Verhoeft, Roelofs & Chwilla, 2009). However, although inhibition is a possible mechanism for language control and is further supported by the L1 global slowing effect, other models have been proposed that explain language control without taking inhibition into account (e.g., Costa, Miozzo & Caramazza, 1999; Finkbeiner, Almeida, Janssen & Caramazza, 2006; La Heij, 2005; Roelofs, 1998) or by suggesting an interaction between inhibition and activation (Grainger & Dijkstra, 1992).

A major novel finding of the present study is that switch costs could be obtained in a handwritten language production task. This has theoretical and practical implications. Although some comprehensive models have been suggested for literacy in bilinguals (e.g., Li, Koh, Geva, Joshi & Chen, 2020), these models focused on word decoding and comprehension as important

cognitive factors for reading comprehension. As our results show that bilinguals use control processes during writing, models about biliteracy may benefit from incorporating mechanisms of language control, particularly for writing.

More detailed processing steps have been postulated in cognitive models of reading, such as the dual-route model and the connectionist model (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Seidenberg & McClelland, 1989), and similar models have been proposed for spelling (e.g., Houghton & Zorzi, 2003). However, as the input in these models are words (either written or spoken), they are not directly applicable to the current experiment. Instead, models of written word naming (e.g., Bonin et al., 2012) that start with pictures as input are more appropriate. Bonin et al. (2012) suggested a limited-cascading model for writing in which information cascades from the semantic to the orthographic L-levels but not from the object recognition to semantic levels. The finding of switch costs for writing output in our study suggests that a mechanism of language control should be added to the model by Bonin et al. (2012), if the model is applied to written word production in bilinguals. Our results, however, do not allow a conclusion about the locus of language control. Whether control occurs within or outside the language system or whether it occurs at a particular level within the language system are questions that are more general. They are also discussed in the area of spoken word production in bilinguals (e.g., Declerck & Philipp, 2015) and require more research to be clarified.

In addition to theoretical implications about control mechanisms in bilingual writing, the finding of switch costs with written output mode has some practical implications. It opens new possibilities to apply this paradigm to research in areas where speaking or audio recordings are not possible or where the primary form of communication is in written rather than in spoken form. Future studies may use handwriting to investigate language control in specific groups such as people with dyslexia, or research topics which are related to the processing of writing but not speaking.

Apart from the output modality, our experiment also manipulated the input modality. In a previous study comparing visual and auditory inputs with vocal output in language switching, Declerck et al. (2015b) found that switch costs in RT were higher with visual than auditory stimuli. Although the present study corroborates the results from Declerck et al. (2015b) that switch costs could be obtained with auditory stimuli, the opposite result was found according to which switch costs in RT were higher with auditory than visual stimuli. The opposite directions of the switch cost differences between auditory and visual stimuli warrant a more detailed discussion. Declerck et al. (2015b) suggested three potential explanations of their results related to the duration of lexical-semantic processing, auditory-vocal interference, and sensory-motor compatibility, which we will discuss in the light of our results.

As the main explanation of their results, Declerck et al. (2015b) argued that lexical-semantic processing takes longer with auditory than visual stimuli, allowing for enhanced language control. They proposed that this implicitly leads to a longer ISI, which has two effects, both leading to lower switch costs with auditory than visual stimuli: longer preparation time and a potential decay effect. When considering whether this explanation can account for the results in the current study, the timing differences between the two studies need to be taken into account. While Declerck et al. (2015b) kept the time from response registration (presumably

end of response) to subsequent stimulus onset constant, thereby letting ISI vary with longer RTs after auditory than visual stimuli, in our experiment we kept ISI constant and let response (onset) to stimulus interval (RSI) vary. Moreover, stimulus duration (3 seconds) and SOA (5 seconds) were rather long in our experiment, but both parameters were shorter in the Declerck et al. (2015b) study. The different stimulus durations and SOAs in the two studies may have induced differences in preparation time. As suggested by Declerck et al. (2015b), participants started preparing during the ISI in their study, which might be caused by the shorter SOA and the early disappearance of the stimulus. Given that auditory stimuli resulted in longer ISI, participants spent more time preparing in the auditory condition, which resulted in reduced switch costs for auditory stimuli. In contrast, in our experiment, where SOA was long and the stimuli still being presented while participants responded, participants might have started preparing for the next trial later, only after finishing the response. RSI and time after response completion were longer for visual than auditory conditions in our experiment, which might have led to longer preparation time for visual than auditory stimuli and therefore reduced switch costs for visual stimuli. Thus, the size of the switch cost may depend on preparation time in both experiments, but preparation time can vary between stimulus modalities depending on experimental parameters.

As a second consequence of longer ISI for auditory stimuli, Declerck et al. (2015b) suggested that a larger decay of language activation from previous trials would occur for auditory compared to visual stimuli, thereby decreasing switch costs. In our study, such an effect is not expected, as ISI was constant for auditory and visual modalities.

As an alternative explanation, Declerck et al. (2015b) suggested that hearing one's own voice in the previous trial could facilitate production in repetition trials or cause interference in switch trials. Accordingly, lower switch costs with auditory stimuli would arise, because the auditory stimuli could overwrite the memory of the previous vocal response (and thereby reduce facilitation and interference), but the visual stimuli could not. Given that this situation was the same in the vocal conditions in our experiment, but that the modality effect was reversed, the explanation seems unlikely to be true.

Declerck et al. (2015b) also mentioned sensory-motor modality compatibility as a third explanation that potentially could explain their results, but which they considered unlikely given the blocked presentation. Our results did not find any interaction effect between input and output modality on switch cost. We therefore agree with Declerck et al. (2015b) that modality compatibility is an unlikely factor to explain their results.

The absence of a switch cost interaction between input and output modalities is different from the finding by a previous study related to task switching, which has shown lower switch costs in modality compatible tasks than incompatible tasks (Stephan & Koch, 2010). The main reason for the absence of an interaction in our study may be that the modalities were blocked in our experiment, but varied within blocks in the study by Stephan and Koch (2010). Modality compatibility effects on switch cost may require variability in input and output modalities in the same block (Fintor, Stephan & Koch, 2018).

Another potential reason for the absence of the modality compatibility effect is the long RSI used in the current study. In the study by Stephan and Koch (2010), the effect of modalities on switch costs were found in the short RSI condition (600 ms), but not in the long RSI condition (1600 ms), because the modality

compatibility effect was due to a short-lived priming component. Nevertheless, in the present study, the duration of stimulus presentation was fixed to three seconds, so the RSI was quite long (around 4 seconds), and it may explain the absence of interactions on switch costs between modalities. However, a long RSI was inevitable in the present study since it was necessary to ensure adequate time for participants to finish writing the whole word in the manual conditions. Moreover, the duration of stimulus presentation had to be the same in the vocal conditions to allow fair comparisons on RTs between output modalities.

Moreover, the absence of a switch cost interaction between modalities may be attributed to the different experiences on language control and multi-tasking. According to our questionnaire, bilinguals switched their language frequently in daily conversation and texting, showing that non-target language inhibition was common. More importantly, bilinguals do not only inhibit non-target language during language switching, but also in every single-language context. Consequently, bilinguals should have more experience on language control than task switching, leading to a higher efficiency on language control. This echoes the finding by Weissberger et al. (2015) that bilinguals had higher efficiency on sustaining inhibition of a non-target language than a non-target task. Due to the different experiences on language control and task switching, the results of task switching (e.g., Stephan & Koch, 2010, 2011) were not able to generalize to the findings in the current study.

Besides switch costs, we found that RT was also affected by input and output modalities. Longer overall RTs were found in the auditory conditions, compared with the visual conditions, in both the present study and the previous study by Declerck et al. (2015b). An explanation proposed by Declerck et al. (2015b) was that the lexical-semantic processing was longer with auditory than visual stimuli. Alternatively, the visual stimuli in the experiment were static, while the auditory stimuli changed over time, which might have led to a perceptual delay in stimulus processing until lexical-semantic processing was engaged.

Overall RTs were longer in the manual than the vocal conditions, and it can be explained in several ways. Writing encompasses more intricate coordination of language-specific and motor control processes than speaking, so it may take longer time to produce (Bonin et al., 2012). This is exacerbated by hand-eye coordination in writing, which may further delay handwriting onset (Perret & Laganaro, 2013). Alternatively, while speaking also requires the coordination of motor control processes of the vocal tract, this skill is learned at a younger age and practiced more frequently than writing, which may lead to a higher automaticity and faster RT in speaking than writing. However, we cannot completely exclude the possibility that our experimental setup also influenced RT differences between speaking and writing. Participants might have applied different strategies for keeping track of the language switching sequence by relying on their memory in the vocal conditions, but by deriving the switching sequence from the previously written words that were still visible on paper in the manual conditions.

Finally, several potential limitations should be noted. First of all, the number of trials per condition was lower in our study than in some previous studies (e.g., Declerck et al., 2015b). To compensate for a loss of statistical power, we increased the sample size from 36, as estimated by a power analysis for detecting two-way interactions with a medium effect size, to 48. While we assume that the power in our experiment was sufficient to detect 2-way interactions with medium effect sizes, power might be

somewhat limited to detect effects of three-way interactions, and may be insufficient to detect a four-way interaction. Therefore, the interpretation of the four-way interaction and to some degree of the three-way interactions needs to be cautious. However, the main findings of our study (*input modality x language transition* and *output modality x language*) are two-way interactions for which power should be sufficient.

Another potential limitation of the present study is related to the possibility of categorical effects of our stimuli. As both images and sound had to represent a same concept, the stimulus choice was limited. Therefore, our stimuli included five concepts which were animals, two of them were instruments and the remaining three did not belong to a clear category (see Appendix for the stimulus list). In fact, category repetition affected RT and accuracy in our data (not reported), but it is unlikely that this effect modulated switch costs, as our stimuli appeared equally in repetition and switch trials in each condition. Nonetheless, future studies may consider avoid using stimuli belonging to a same category to rule out the category effect.

A further potential limitation might be that participants were required to respond while the stimulus was still being presented. Potentially, participants could have been hesitant to speak when they were listening to the sound as listening and speaking seldom happen together. This issue seems to be related to the differences from the study by Declerck et al. (2015b) that we discussed in detail above. Future studies would be necessary to investigate such modality differences to test at which level of cognitive processing they occur and how they are modulated by experimental parameters, such as differences in timing.

Furthermore, regarding the L1 global inhibition effect, a baseline single-language condition was not included. In other words, one may argue that even for monolinguals, the RT of traditional Chinese writing may be slower than that of English writing (for example, due to different and potentially higher motor demands for writing Chinese compared to English), so the L1 slowing just reflects baseline difference, but not global inhibition. We believe that this possibility is low because in the present study participants were more proficient in Chinese than English and the frequency of the words in both languages was high. Moreover, the RT was defined as the duration between the onset of stimulus appearance and the first contact between the pen and the paper, and the writing time was not counted. Therefore, there should be no time difference in the semantic, orthographic and grapheme processing before writing between two languages. However, future research may consider including a baseline single-language condition to measure the reaction time of writing in different languages in monolinguals when the L1 global inhibition effect is examined.

A final potential limitation relates to the generalization to other types of bilinguals. Participants in the present study were early bilinguals as they had started learning English before three years old on average. However, due to the significant difference between English and Chinese proficiency (see table 1), they cannot be considered as balanced bilinguals. As previous studies have pointed out that language proficiency affects switch cost (e.g., Meuter & Allport, 1999), our current results may not be able to generalize to balanced bilinguals, and future research studies are needed to investigate the writing switch cost in balanced bilinguals. Furthermore, the scripts of Chinese and English are morphosyllabic and alphabetic respectively, which are very different. Hence, whether the switch costs in writing can be found in other more similar languages is unknown and it warrants future research.

In conclusion, the present study showed that switch costs in terms of error percentage and RT did not differ between speaking and writing, when bilingual participants switched between Chinese and English. The switch cost findings have shown that the language control mechanism may be similar in both spoken and written word production, and they may both rely on the inhibition of the non-target language in bilinguals. In addition, the results showed that switch costs differ between auditory and visual stimuli, potentially driven by differential preparation. However, the different direction compared to a previous study suggests that experimental parameters may induce certain preparatory processes. Finally, the existence of switch costs in handwriting is a novel finding. It may contribute to the reading and writing models in bilinguals by suggesting the need for language control. Moreover, it may provide a new methodology for future research related to language switching with written word production.

Competing interests. The authors declare none

Acknowledgements. We would like to thank the reviewers for their constructive comments on previous versions of the manuscript.

Notes

ⁱ “L-level” is defined as a lexical representation retrieved from long-term memory (Goldrick & Rapp, 2007), and it was added to the model to remain neutral in a debate about the existence of lemma level between the semantic representation and phonological stage (Levelt, 1989; Caramazza, 1997).

ⁱⁱ We conducted an a priori power analysis with G*Power (Faul, Erdfelder, Lang & Buchner, 2007) to determine the required sample size to observe a statistically significant result in the current experiment. Given an alpha of .05 and an effect size based on a study by Declerck et al., 2015b (Cohen’s $f = .25$), the required sample size of the two-way interaction was $N = 36$. However, since there are 48 combinations for a full counterbalancing of all four modalities and two language switching sequences, the final sample size is $N = 48$.

References

- Bobb SC and 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**, 568–585. doi: 10.1080/20445911.2013.792822
- Boersma P and Weenink D** (2019) Praat: doing phonetics by computer [Computer program]. Version 6.0.46, retrieved 24 January 2019 from <http://www.praat.org/>
- Bonin P, Peereman R and Fayol M** (2001) Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language* **45**, 688–720. doi: 10.1006/jmla.2000.2786
- Bonin P, Roux S, Barry C and Canell L** (2012) Evidence for a limited-cascading account of written word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **38**, 1741–1758. doi: 10.1037/a0028471
- Brooks LR** (1968) Spatial and vocal components of the act of recall. *Canadian Journal of Psychology/Revue Canadienne de Psychologie* **22**, 349–368. doi: 10.1037/h0082775
- Caramazza A** (1997) How many levels of processing are there in lexical access? *Cognitive Neuropsychology* **14**, 177–208. doi: 10.1080/026432997381664
- Christoffels IK, Firk C and Schiller NO** (2007) Bilingual language control: An event-related brain potential study. *Brain Research* **1147**, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Coltheart M, Rastle K, Perry C, Langdon R and Ziegler J** (2001) DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* **108**, 204. doi: 10.1037/0033-295x.108.1.204
- Costa A and 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**, 491–511. doi: 10.1016/j.jml.2004.02.002
- Costa A, Miozzo M and Caramazza A** (1999) Lexical selection in bilinguals: Do words in the bilingual’s two lexicons compete for selection?. *Journal of Memory and Language* **41**, 365–397. doi: 10.1006/jmla.1999.2651
- Costa A, Santesteban M and 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**, 1057–1074. doi: 10.1037/0278-7393.32.5.1057
- Declerck M and Philipp AM** (2015) A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review* **22**, 1630–1645. doi: 10.3758/s13423-015-0836-1
- Declerck M, Koch I and Philipp AM** (2015a) The minimum requirements of language control: Evidence from sequential predictability effects in language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 377–394. doi: 10.1037/xlm0000021
- Declerck M, Stephan DN, Koch I and Philipp AM** (2015b) The other modality: Auditory stimuli in language switching. *Journal of Cognitive Psychology* **27**, 685–691. doi: 10.1080/20445911.2015.1026265
- Faul F, Erdfelder E, Lang A-G and Buchner A** (2007) G*Power 3: A flexible statistical power analysis program for social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**, 175–191. doi: 10.3758/BF03193146
- Finkbeiner M, Almeida J, Janssen N and Caramazza A** (2006) Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 1075–1089. doi: 10.1037/0278-7393.32.5.1075
- Fintor E, Stephan DN and Koch I** (2018) Emerging features of modality mappings in task switching: Modality compatibility requires variability at the level of both stimulus and response modality. *Psychological Research* **82**, 121–133. doi: 10.1007/s00426-017-0875-5
- Geschwind N** (1969) Problems in the anatomical understanding of the aphasias. In Benton AL (Ed.), *Contributions to clinical neuropsychology*. Chicago: Aldine.
- Goldrick M and Rapp B** (2007) Lexical and post-lexical phonological representations in spoken production. *Cognition* **102**, 219–260. doi: 10.1016/j.cognition.2005.12.010
- Gollan TH and Ferreira VS** (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**, 640–665. doi: 10.1037/a0014981
- Grainger J and Dijkstra T** (1992) On the representation and use of language information in bilinguals. In Harris RJ (Ed.), *Cognitive processing in bilinguals*. Amsterdam, The Netherlands: North Holland.
- Green DW** (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* **1**, 67–81. doi: 10.1017/S1366728998000133
- Holcomb PJ and Neville HJ** (1990) Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, **5**, 281–312. doi: 10.1080/01690969008407065
- Houghton G and Zorzi M** (2003) Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology* **20**, 115–162. doi: 10.1080/02643290242000871
- Kaufmann E, Mittelberg I, Koch I and Philipp AM** (2018) Modality effects in language switching: Evidence for a bimodal advantage. *Bilingualism: Language and Cognition* **21**, 243–250. doi: 10.1017/S136672891600122X
- La Heij W** (2005) Selection processes in monolingual and bilingual lexical access. In Kroll JF & de Groot AMB (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford, England: Oxford University Press.
- Leblanc R and Painchaud G** (1985) Self-assessment as a second-language placement instrument. *TESOL Quarterly* **19**, 673–686. doi: 10.2307/3586670
- Levelt WJ** (1989) *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Li M, Koh PW, Geva E, Joshi R and Chen X** (2020) The componential model of reading in bilingual learners. *Journal of Educational Psychology*. doi: 10.1037/edu0000459

- Logie RH, Zucco GM and Baddeley AD** (1990) Interference with visual short-term memory. *Acta Psychologica* 75, 55–74. doi: 10.1016/0001-6918(90)90066-O
- Luria AR** (1970) *Traumatic aphasia. La Haye* The Netherlands: Mouton de Gruyter.
- Meuter RF and Allport A** (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40, 25–40. doi: 10.1006/jmla.1998.2602
- Perret C and Laganaro M** (2012) Comparison of electrophysiological correlates of writing and speaking: a topographic ERP analysis. *Brain Topography* 25, 64–72. doi: 10.1007/s10548-011-0200-3
- Perret C and Laganaro M** (2013) Why are written picture naming latencies (not) longer than spoken naming? *Reading and Writing* 26, 225–239. doi: 10.3389/fpsyg.2016.00031
- Pinet S, Ziegler JC and Alario FX** (2016) Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin & Review* 23, 1898–1906. doi: 10.3758/s13423-016-1044-3
- Prior A and Gollan TH** (2011) Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society* 17, 682–691. doi: 10.1017/S1355617711000580
- Roelofs A** (1998) Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition* 1, 94–95. doi:10.1017/S1366728998000194
- Schaeffner S, Fibla L and Philipp AM** (2017) Bimodal language switching: New insights from signing and typing. *Journal of Memory and Language* 94, 1–14. doi: 10.1016/j.jml.2016.11.002
- Seidenberg MS and McClelland JL** (1989) A distributed, developmental model of word recognition and naming. *Psychological review* 96, 523. doi: 10.1037/0033-295x.96.4.523
- Snodgrass JG and Vanderwart M** (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* 6, 174–215. doi: 10.1037//0278-7393.6.2.174
- Stephan DN and Koch I** (2010) Central cross-talk in task switching: Evidence from manipulating input–output modality compatibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 1075–1081. doi: 10.1037/a0019695
- Stephan DN and Koch I** (2011) The role of input–output modality compatibility in task switching. *Psychological Research* 75, 491–498. doi: 10.1007/s00426-011-0353-4
- Thomas MS and Allport A** (2000) Language switching costs in bilingual visual word recognition. *Journal of Memory and Language* 43, 44–66. doi: 10.1006/jmla.1999.2700
- Verhoef KMW, Roelofs A and Chwilla DJ** (2009) Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience* 22, 1832–1843. doi: 10.1162/jocn.2009.21291
- Weissberger GH, Gollan TH, Bondi MW, Clark LR and Wierenga CE** (2015) Language and task switching in the bilingual brain: Bilinguals are staying, not switching, experts. *Neuropsychologia* 66, 193–203. doi: 10.1016/j.neuropsychologia.2014.10.037
- Yau MS** (1993) Functions of two codes in Hong Kong Chinese. *World Englishes* 12, 25–33. doi: 10.1111/j.1467-971X.1993.tb00004.x

Appendix. Ten concepts used in the experiment

Traditional Chinese / Cantonese	English
車	Car
貓	Cat
牛	Cow
狗	Dog
門	Door
鼓	Drum
笛	Flute
槍	Gun
馬	Horse
羊	Sheep