Bilingualism: Language and Cognition

cambridge.org/bil

The role of audiovisual modality in predicting the neurodynamics of language control in Tibetan–Chinese bilinguals

Yanbing Hu[1,2,3]*, Keyu Pan, [1,2,3]* and Xiaofeng Ma[1,2,3]

[1]Department of Psychology, Northwest Normal University, Lanzhou, Gansu, China; [2]Key Laboratory of Education Digitalization of Gansu Province and [3]Key Laboratory of Behavior and Mental Health of Gansu Province

Research Article

Cite this article: Hu, Y., Pan, K., and Ma, X. (2025). The role of audiovisual modality in predicting the neurodynamics of language control in Tibetan–Chinese bilinguals. *Bilingualism: Language and Cognition* 1–13. https://doi.org/10.1017/S1366728925100424

Received: 23 December 2024
Revised: 2 July 2025
Accepted: 24 July 2025

Keywords:
audiovisual modality; language switching; Tibetan–Chinese bilinguals; ERP; machine learning

Corresponding author:
Xiaofeng Ma;
Email: Maxiaofeng@nwnu.edu.cn

*Yanbing Hu and Keyu Pan contributed equally to the work.

Abstract

Although bilinguals use both auditory and visual cues, the cognitive cost of language switching in audiovisual contexts is unclear. We investigated the cost in Tibetan–Chinese bilinguals using a task with audiovisual, visual and auditory modalities. In Study 1, the audiovisual modality yielded the fastest reaction times, reflecting improved processing efficiency. ERP data revealed smaller positive amplitudes in the early window (200–350 ms) for audiovisual modality, indicating reduced neural demand, while only auditory modality showed significant divergence in the later window (350–700 ms). Moreover, audiovisual context, L2-to-L1 switching and early neural responses predicted switching behavior. Study 2 replicated the behavioral and ERP findings of Study 1 and demonstrated that auditory input and second-language processing exacerbated switch costs. These findings shed light on multisensory integration in language switching by demonstrating that audiovisual cues reduce switch costs, whereas auditory input and second-language processing exacerbate them, with implications for language education and cognitive interventions.

1. Introduction

Bilinguals frequently switch between languages during daily communication, which is a common phenomenon in multilingual environments (Seitz & Smith, 2022). Language switching refers to the ability of bilinguals or multilinguals to seamlessly alternate between two or more languages depending on the context, or to use multiple languages within a single conversation (Gullifer & Titone, 2020; Smith et al., 2020). This process involves complex cognitive control, where bilinguals need to flexibly switch between different language systems while maintaining fluent communication and avoiding confusion (Kheder & Kaan, 2021). However, real-life communication is often multimodal, involving the simultaneous processing of various sensory inputs, such as visual and auditory information. This adds another layer of complexity to the language switching control process. In certain regions of China, people not only need to master their mother tongue but also use a second language (L2) in their daily lives. For example, Tibetan university students often switch between Tibetan and Mandarin in daily communication, which further complicates the language control demands in such multilingual environments. This study aims to explore the neural dynamics of language switching in Tibetan–Mandarin bilinguals in different sensory modalities, such as visual, auditory and audiovisual conditions. By analyzing the language switching control mechanisms in bilinguals across these multimodal conditions, the research seeks to uncover how sensory modalities influence the neural processing during language switching. Unlike earlier studies on Indo-European (e.g., Spanish–English) or cross-linguistic bilingual switching (e.g., Mandarin–English) (Flege & Bohn, 2021), this study examines Tibetan native speakers (Sino-Tibetan languages). Driven by Tibetan students' growing reliance on Mandarin for education, work and social life (Gao & Zeng, 2021), exploring Tibetan–Mandarin switching deepens our understanding of minority–national language interactions and their influencing factors.

2. Literature review

2.1. Bilingual language switching from the perspective of the inhibitory control model (ICM)

The ICM is an important theoretical framework for explaining the cognitive processes inlying bilingual language switching, particularly focusing on the difficulty associated with switching (i.e., switch costs) (Declerck et al., 2015; D. W. Green, 1998). The model introduces the concept of "sustained inhibition": when a bilingual uses a specific language in one task (task $n-1$), the other language is inhibited. If the subsequent task (task n) requires switching to the previously inhibited language, overcoming that inhibition adds difficulty to the switch (Linck et al., 2012). Conversely,

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

if the same language is used in consecutive tasks (i.e., a repeat task), there is no need to overcome inhibition, resulting in smoother processing. Thus, this model explains why language switching is typically more effortful and time-consuming compared to repeating the same language (Goldrick & Gollan, 2023). For instance, if a Tibetan–Chinese bilingual describes a picture in Tibetan during the first task, Mandarin activation is inhibited at that point. If the subsequent task requires switching to Mandarin to describe another picture, the individual must overcome the inhibition of Mandarin, resulting in higher switch costs (Costa & Santesteban, 2004; Schwieter & Sunderman, 2008). Schwieter and Sunderman's (2008) Proficiency-Specific Model posits that second language learners exhibit language control processes similar to those described in ICM, while highly proficient bilinguals differ (Costa & Santesteban, 2004). According to this model, highly proficient bilinguals rely on language cues that, at the conceptual stage, locate the target language (La Heij, 2005). These language cues ensure that more activation flows to the representation of the target language, rather than to nontarget language representations (e.g., translation equivalents). As a result, highly proficient bilinguals do not need to inhibit between different language representations. This assumption is primarily based on the lack of asymmetrical switch costs in highly proficient bilinguals during language switching, meaning that the cost of switching from L2 to L1 does not differ from that of switching from L1 to L2 (Calabria et al., 2011; Christoffels et al., 2007; Costa et al., 2006; Costa & Santesteban, 2004).

Symmetrical switch costs were first observed by Meuter and Allport (1999), indicating that language proficiency significantly affects switch costs. By categorizing participants into highly proficient bilinguals and second language learners, they found that the highly proficient group exhibited symmetrical switch costs, while second language learners displayed asymmetrical switch costs. Research by Filippi et al. (2014) also supports this conclusion, showing that asymmetrical switch costs in naming tasks are negatively correlated with L2 proficiency. Furthermore, Schwieter and Sinman (2008) found that switch costs became symmetrical once a certain L2 proficiency threshold was reached (as measured by language fluency). These studies suggest that language proficiency plays a crucial role in language switching, particularly in explaining asymmetrical switch costs. The symmetrical switch costs observed between two highly proficient languages can be explained by reactive and sustained inhibition. This implies that, due to the similar activation levels of both languages, the degree of inhibition is also similar, resulting in equivalent amounts of inhibition to overcome during switching tasks, and consequently leading to similar cross-language switch costs.

2.2. The influence of modalities on bilingual language switching

Compared to previous studies that mainly investigated bilingual switching through visual stimuli, some research has incorporated both visual and auditory cues to explore the cost of language switching among bilinguals (Declerck et al., 2015). The results showed that participants exhibited switch costs in both visual and auditory conditions, with the switch cost being significantly higher in the visual condition than in the auditory condition (Declerck et al., 2015). This may be due to the longer processing time required for auditory information during language switching, coupled with the predictability of language switching sequences. This extended processing time may actually facilitate switch preparation, thereby reducing the switch cost (Declerck et al., 2015). The study highlighted differences in bilingual switch costs across different modalities and proposed a new paradigm for studying switch costs through the auditory modality stimulus. However, this study did not delve deeply into whether this result was due to the experimental manipulation itself or was influenced by participants' expectation or anticipation effects during the experiment. Additionally, other research further examined Cantonese–Mandarin bilinguals performing language switching tasks in visual and auditory modalities. Results showed that, compared to the visual modality stimulus, reaction times for switching tasks in the auditory modality stimulus were significantly longer (Xing Qiang, 2021). Unlike the study by Declerck et al. (2015), the target language sequence in the study referenced as Xing Qiang (2021) was unpredictable, suggesting that in such an unpredictable context, the auditory modality stimulus may not facilitate language switching for bilinguals. This suggests that processing the second language in the auditory modality (versus other modalities) may produce larger differences between switching and repeating tasks.

However, previous research indicates inconsistencies regarding the influence of different modality types on bilingual switch costs. Few studies have explored how audiovisual modality (as opposed to single modalities) stimuli affect bilingual switching. In real-life communication, the environment is typically multimodal, where people rely on various sensory inputs, such as visual and auditory information, simultaneously. This multisensory interaction not only makes communication richer and more flexible but also enhances the brain's ability to process and integrate multiple types of sensory information, improving language comprehension and expression efficiency. More importantly, increasing evidence suggests that audiovisual modality stimuli play a crucial role in second language (L2) acquisition, particularly in vocabulary and grammar learning (Muñoz et al., 2023). Some studies have shown that combining visual and auditory input can strengthen language learners' instanding and memory of language structures, especially when learners watch audiovisual materials with subtitles or when spoken words and text appear in sync (Muñoz et al., 2023). This approach helps deepen learners' retention of L2 vocabulary and improves their grasp of grammatical rules (Çekiç, 2024). The reason audiovisual modality stimulus may enhance second language (L2) learning is that lexical access within the language system involves the activation of multimodal information (Athanasopoulos et al., 2015; Wang & Wei, 2023). For instance, when participants see an image of a bird while hearing the corresponding bird call, the activation of target language vocabulary is likely to become more automatic (Asaadi et al., 2024). According to the ICM, bilinguals inhibit the activation of one language system while using the other, thereby reducing interference. However, audiovisual input provides stronger contextual support and sensory cues, facilitating faster language switching and activation of target language vocabulary, which reduces inhibition and enhances fluency and flexibility (Declerck et al., 2015). For example, speakers of English tend to produce fewer manner-related gestures than speakers of Chinese (Morett et al., 2022).

In summary, audiovisual input not only enhances learners' language skills but also offers additional contextual cues, enabling them to use the language more flexibly in real-life communication. Building on these insights, the current study aims to investigate this further by focusing on the mechanisms through which an audiovisual modality stimulus influences bilingual language switching.

2.3. Neural dynamics evidence of bilingual language switching

EEG technology is highly suitable for detecting the fine temporal granularity of brain neural dynamics. Unlike behavioral indicators, which can only reflect the outcome of bilingual switching costs

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

(such as reaction times), neural signals can more directly capture the real-time brain activity during bilingual switching, providing more precise and immediate data on cognitive processes (Pereira Soares et al., 2024). In an experiment by Lavric et al. (2019), participants were asked to view black-and-white pictures of everyday objects and name them, with the language cue provided via spoken instructions (e.g., “Deutsch” or “English”) or accelerated segments of the German or British national anthems. The language cues were alternated across trials, with random switches in some trials. The experimental results showed that in the 300–500 ms and 500–700 ms time windows, the switching condition (compared to the repeat condition) elicited larger positive ERP amplitudes. However, Lavric et al.’s (2019) findings were not supported by the language-switching ERP study conducted by Verhoef et al. (2010). In their study, Verhoef et al. (2010) found that in the 200–350 ms window, language cues in switching trials elicited more negative amplitudes at frontal electrodes compared to repeat trials. A study employed a language switching and head position task, in which participants performed a joint naming–listening task (Liu et al., 2024). One participant (Participant A) named pictures, while the other participant (Participant B) orally identified the head of the compound word after hearing Participant A’s naming (Liu et al., 2024). The language for naming was indicated by color cues: red for L1 (Chinese) and blue for L2 (English). The ERP results revealed an N2 polarity reversal effect between Participants A and B. Specifically, in Participant A’s neural signals, a more negative wave was elicited during switch trials compared to nonswitch trials (similar to the findings of Verhoef et al. (2010)). In contrast, Participant B’s neural signals showed a more positive wave for switch trials compared to nonswitch trials (similar to Lavric et al. (2019)). Some research has suggested that the findings of Lavric et al. (2019) and Verhoef et al. (2010) on bilingual switching ERP components may not be contradictory (Declerck et al., 2021). This is due to an earlier task-switching study in which Lavric et al. (2008) identified a dipole ERP component associated with cue processing, which included a frontal switching negativity and a posterior switching positivity, both operating concurrently and interdependently. Therefore, the posterior switching positivity observed by Lavric et al. (2019) and the frontal switching negativity observed by Verhoef et al. (2010) may represent two aspects of the same dipole ERP component. Additionally, other studies have found dipole characteristics in the late window: in the 350–700 ms time window following stimulus presentation, switching trials (compared to repeat trials) elicited a stronger posterior positivity (LPC) and a larger frontal negativity. These results reflect the late dipole component characteristics during language switching, revealing differences in regional neural activity during the switching task (Declerck et al., 2021). However, all of these studies only used the visual modality stimulus to examine the neural dynamics of language switching in bilinguals. To further clarify how different modalities influence the cognitive processes involved in bilingual language switching at early (200–350 ms) and late (350–700 ms) stages, this study will compare the effects of auditory, visual and audiovisual stimuli on this process.

2.4. Neurodynamic signals in the bilingual brain predict behavioral performance

Furthermore, the current study continues to focus on elucidating the relationship between the neurodynamic signals of Tibetan–Chinese bilinguals and their behavioral performance. Clarifying this issue not only helps to deepen our instanding of how the brain collaborates with second language behavior, but also provides key neurophysiological evidence for this process. These findings will offer solid empirical support for further uncovering the cognitive control mechanisms, language switching strategies and neural basis involved in bilingual language use, thereby contributing to a more comprehensive instanding of brain function in bilinguals.

However, previous research on language switching has rarely linked neural signals with behavioral performance (Peeters, 2020), which limits the understanding of the cognitive processes underlying ERP signals. Additionally, most previous studies have used linear regression methods that typically focus on the predictive power of a single neural signal on behavior (Ou & Law, 2017). Although this approach provides insights into the linear associations between neural and behavioral aspects to some extent, its limitation lies in its inability to fully capture the interactive effects of multidimensional signals in complex brain activities or determine which neural signals play a more significant role in predicting behavioral performance. To overcome this limitation, the present study incorporates a Ridge regression machine learning model (Hu et al., 2024). Compared to traditional linear regression methods that consider only a single variable, machine learning models possess a stronger capability to handle multivariable and multifeature data, enabling the integration of multiple neural signal features to reveal more complex, nonlinear relationships between neurodynamics and behavior (Rajan, 2022).

Ridge regression is particularly suitable for addressing multicollinearity, avoiding overreliance on any specific variable (Venkatesh et al., 2023). By applying the Ridge regression model, this study aims not only to more accurately predict behavioral performance but also to enhance our instanding of how multiple neural activity signals collectively influence bilinguals during language switching and processing. Through this approach, we hope to provide deeper insights into the neural mechanisms of Tibetan–Chinese bilinguals during language switching, offering new perspectives and richer empirical evidence for bilingual cognitive neuroscience. This instanding will not only contribute to the development of bilingual education and language policies but may also provide theoretical support for the diagnosis and intervention of language disorders. Specifically, our findings are expected to provide valuable insights for the development of bilingual education and language policies. For example, by identifying neural markers associated with efficient language switching, our results can help educators and policymakers understand which neural mechanisms are closely linked to language switching performance, thereby guiding the formulation of more targeted teaching methods and curriculum designs. Furthermore, through an in-depth analysis of the neural mechanisms underlying language control, our findings may also offer theoretical support for the diagnosis and intervention strategies for language disorders. In particular, for language disorders commonly observed in bilingual populations (such as developmental language disorder), these neural markers could contribute to refining diagnostic criteria and establishing more effective intervention measures.

2.5. The current study

Previous research has predominantly focused on the inhibitory mechanisms and switch costs involved in bilingual language switching. However, these studies have largely been limited to experimental designs employing unimodal stimuli, lacking a comprehensive examination of how audiovisual information processing affects language control. The present study aims to investigate the cognitive and neurodynamic characteristics of language switching in Tibetan–Chinese bilinguals in different modalities, including visual, auditory and audiovisual conditions. Based on previous research, the study further seeks to determine which neural dynamic signals elicited by

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

different sensory modalities during language switching exhibit greater predictive value for L1–L2 switching behavior. Specifically, a cue-based language switching task is employed in which participants are required to name images in the target language indicated by different colored cues, while both electroencephalogram (EEG) data and reaction times are recorded throughout the task.

Based on previous research, the following hypotheses are proposed:

H1: Given that the participants are Tibetan–Chinese university students, who are proficient in their second language (Mandarin), no significant asymmetrical switch costs are expected between the different language switching directions (L1 → L2 versus L2 → L1).

H2: In unimodal conditions, it is expected that switch costs in the auditory modality stimulus will be higher than those in the visual modality stimulus, reflecting the modulatory effect of different sensory modalities on language switching. This switching cost is particularly pronounced when processing the second language.

H3a: If switch costs in the audiovisual modality stimulus are lower than those in the unimodal conditions (visual and auditory), this would indicate that multimodal integration helps to reduce the language switching load associated with language switching.

H3b: If switch costs in the audiovisual modality stimulus are higher than or equal to those in the unimodal conditions, this may suggest that audiovisual stimuli are redundant in the context of language switching.

3. Study 1

3.1. Method

3.1.1. Participants

The current study recruited 21 university students proficient in both Tibetan and Mandarin (5 males), with an average age of approximately 20.69 years (SD = 0.9). The participants are native speakers of Tibetan (L1) and began systematic learning of Mandarin (L2) after kindergarten, around the age of 3. They self-reported high proficiency in both Tibetan and Mandarin (the current study used a 7-point Likert scale for assessment, where 1 indicates very low proficiency and 7 indicates very high proficiency), with no significant difference in proficiency between the two languages (t (20) = 0.49, p > .05; see Table 1 for additional comparisons of language abilities). Additionally, they all had normal vision and hearing and no history of psychiatric or neurological disorders. None of the participants had participated in similar experiments before, and they volunteered to take part in this study. Upon completion of the experiment, they will receive appropriate compensation. We estimated the minimum sample size required to achieve over 95% power with an α level of 0.05. Using the *powerCurve* function, with a set seed of 123 and 1000 simulations, the power analysis indicated that at least 16 participants are needed to achieve a statistical power of over 95% to detect the given effect size (P. Green & MacLeod, 2016; Hu et al., 2023). Additionally, a power analysis was conducted using G*Power 3.1.9.2 with the following settings: F-test for repeated measures ANOVA (within factors), effect size $f = 0.25$ (medium effect size), α error probability = 0.05, correlation among repeated measures = 0.5, power ($1-\beta$ error probability) = 0.95, one group, 12 measurements and a nonsphericity correction $\varepsilon = 1$ (Faul et al., 2007). The analysis indicated that a minimum sample size of 18 participants is required to achieve the desired power level. Given that the actual number of participants exceeds this estimate, it suggests that the current study's statistical power is adequate.

Table 1. Self-assessment of language proficiency of tibetan–mandarin bilingual participants

Language skill	Tibetan (M ± SD)	Mandarin (M ± SD)	T	p
Reading proficiency	6.71 ± 0.46	6.62 ± 0.50	1.45	.16
Writing	6.43 ± 0.51	6.52 ± 0.51	−1	.33
Speaking ability	6.52 ± 0.51	6.71 ± 0.46	−2.17	.42
Comprehension ability	6.81 ± 0.40	6.71 ± 0.46	1.45	.16

Note: M, mean; SD, standard deviation.

3.1.2. Experimental materials and procedures

The experimental procedure was presented and conducted using E-prime 2.0 software (see Figure 1). Before starting the formal experiment, participants were required to familiarize themselves with the 120 images and 120 audio clips. After familiarization, participants were instructed to name all the displayed images and audio clips to ensure they understood and remembered the names. To ensure participants were proficient with the experimental procedure and could respond accurately as required, they had to achieve at least 95% accuracy in the practice phase before proceeding to the formal experimental phase.

In the visual block, after a "+" was presented for 500 ms, a blank screen appeared for 200 ms, followed by a black-and-white outline image in the center of the screen surrounded by a red or blue frame for 1000 ms. Participants were required to name the image in Tibetan (L1) if the frame was red, and in Mandarin (L2) if the frame was blue.

In the auditory block, after a "+" was presented for 500 ms, a blank screen appeared for 200 ms, followed by an audio clip for 1000 ms. Participants were instructed to name the audio based on the fixation cross color: using Tibetan (L1) when the fixation cross was red and Mandarin (L2) when it was blue, naming the audio as it played.

In the audiovisual block, after a "+" was presented for 500 ms, a blank screen appeared for 200 ms, followed by an image and an audio clip for 1000 ms. The audio matched the semantic content of the image, such as an image of a train accompanied by the sound of a train. The image was surrounded by a red or blue frame for 1000 ms. Participants were required to name the image in Tibetan (L1) if the frame was red, and in Mandarin (L2) if the frame was blue.

It should be noted that, to eliminate any potential differences caused by color cue, we recorded participants' reaction times only after the stimulus was fully presented. Specifically, the sequence for each modality was as follows: a 500 ms fixation, a 200 ms blank screen and a 1000 ms stimulus presentation, with reaction times recorded after a total of 1700 ms.

Within each modality block, there were two types of tasks: a switching task and a repeating task. Specifically, in the switching task, two consecutive images or audio clips required naming in different languages, while in the repeating task, participants were instructed to name the images or audio clips in the same language.

The formal experiment included 560 trials. To minimize familiarity effects with the images and audio, the audiovisual block used

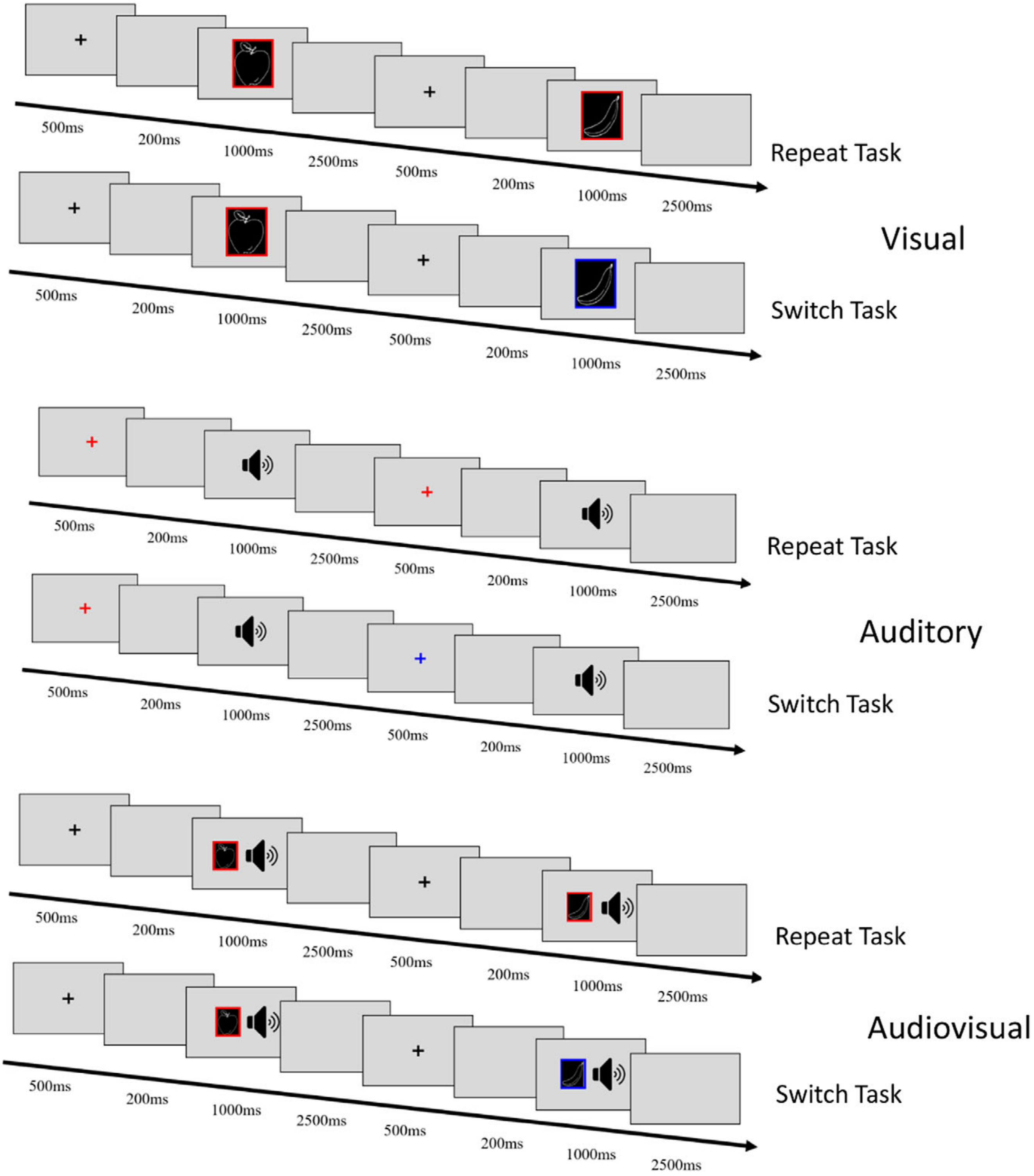

Figure 1. The experimental flowchart illustrates the repeat and switch tasks in visual, auditory and audiovisual conditions. In the diagram, we use the L1 repeat task and the L1–L2 switch task as examples to explain the process. In the L1 repeat task, participants see a red cue, prompting them to name the image in L1. In the L1–L2 switch task, participants first see a red cue, instructing them to respond in L1. Then, in the second trial, they see a blue cue, signaling them to switch to L2 for the response.

different lists of images and audio from the visual block, balanced between participants. Throughout the entire experiment, recording devices were continuously active, and EV Capture software was used for recording to facilitate subsequent analysis of participants' responses. Participants completed the visual, auditory and audiovisual modules, which were balanced across participants.

3.1.3. EEG recording and preprocessing

The EEG data were collected in a dim, sound-attenuated room using a 64-channel electrode cap based on the 10–10 system. During EEG data collection (event-related potentials), we used the ANT system equipped with a 64-channel electrode cap, with FCz as the online reference electrode. During online recording, a

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

band-pass filter was set from 0.1 to 100 Hz, ensuring scalp impedance remained below 10 kΩ, while the sampling rate was maintained at 500 Hz to obtain high-quality EEG signals.

In the data preprocessing stage, to eliminate unnecessary noise, offline filtering was conducted, which included applying a high-pass filter at 0.01 Hz and a low-pass filter at 30 Hz. In the current study, independent component analysis (ICA) was employed to correct ocular artifacts. Correct trials with peak-to-peak deviations exceeding ±70 μV due to artifacts were excluded from the final averaging to ensure data accuracy and reliability. Subsequently, the EEG activity for each condition was averaged, focusing specifically on ERP elicited at the time of stimulus presentation. The analysis window covered 200 ms before stimulus onset (used as a baseline) and 800 ms after stimulus onset to fully capture the dynamic changes in brain activity. Additionally, we averaged the EEG responses associated with the correct responses for each stimulus type.

3.1.4. Data analysis

Behavioral Data Analysis: The current study used mixed linear models for analysis, performed with the *lme4* and *lmerTest* packages in the R environment. Fixed factors included Modality (Auditory/Visual/Audiovisual), Task type (Switching/Repeating) and Language type (Mandarin/Tibetan), while participants were treated as random factors. For all fixed effects, the Satterthwaite approximation was used to assess their significance in the LMM (Luke, 2017). The *confint* function from the *stats* package in R provided confidence intervals for parameter estimation in the LMM. Post hoc comparisons were performed using the emmeans function, with *p*-values adjusted using the Tukey method.

ERP Data Analysis: Based on previous literature on task language switching, we calculated the mean amplitude for each participant within the early time window of 200–350 ms locked to the onset of modality stimuli (Verhoef et al., 2010). We were also interested in the later time window of 350 to 700 ms, also locked to the modality stimulus onset (Lavric et al., 2019). Further, the 200–350 ms and 350–700 ms time windows were treated as representing the N2-like and LPC components, respectively (Lavric et al., 2019; Verhoef et al., 2010). The current study considered Anteriority (Anterior, Central, Posterior) as regions of interest (ROIs). The Anterior region comprised electrodes F1, Fz, F2, FC1, FCz, FC2 and Fpz; the Central region consisted of electrodes C1, Cz, C2, CP1, CPz and CP2; and the Posterior region included electrodes P1, Pz, P2, O1, Oz, POz and O2. Specifically, EEG data analysis was conducted using mixed linear models, performed with the *lme4* and *lmerTest* packages in the R environment. Fixed factors included Modality (auditory/visual/audiovisual), Task type (switching/repeating), Language type (Mandarin/Tibetan) and ROI (anterior, central, posterior), while participants were treated as random factors. The Satterthwaite approximation was used to assess the significance of all fixed effects in the LMM (Luke, 2017). Confidence intervals for parameter estimation in the LMM were provided using the *confint* function from the *stats* package in R. Post hoc comparisons were performed using the *emmeans* function, with *p*-values adjusted using the Tukey method.

Analysis of neural prediction of switching cost: To further investigate how neural signals in the language-switching task predict second-language switching behavior, this study focused on the effect of neural signals related to the switching task on behavioral performance. In the study, data features were derived from neural signals in different conditions during the switching task, and behavioral performance was defined as the "cost," i.e., the average switching cost for transitioning from Tibetan to Chinese in auditory, visual and audiovisual conditions. Each switching cost was calculated by subtracting the reaction time in the repeating task from the reaction time in the switching task. Finally, we combined the costs from the different modalities and used them as the label in the machine learning model. Ridge regression was used for analysis. Compared to univariate linear regression, ridge regression controls model complexity by introducing Tikhonov regularization and has an advantage in handling multicollinearity (Venkatesh et al., 2023).

The study selected 36 feature variables (3 × 2 × 3 × 2), including Modality (Auditory/Visual/Audiovisual), Language type (Mandarin/Tibetan), ROIs (anterior, central, posterior) and two time windows (Early/Late). Specifically, the model includes 36 features and 1 label (the switching behavior data from the combined modalities). In the Python environment, we used the *GridSearchCV* function to perform a grid search to determine optimal parameters. The model was then trained using 1000 iterations of held-out training (80% training set, 20% test set) based on the optimal parameters. To validate the model's effectiveness, a permutation test was conducted by comparing real labels with shuffled labels in 5000 iterations, using mean squared error (MSE) and *R*-squared (R^2) for validation.

3.2. Results

3.2.1. Behavioral results

The current study found a significant main effect of Modality ($F(2, 220) = 126.32$, $p < .001$). Post hoc comparisons revealed that the reaction time for the Audiovisual condition (611.51 ± 393.11 ms) was significantly faster than that for the Auditory condition (1465.00 ± 653.30 ms; $\beta = -853$, SE = 56.5, $t = -15.12$, 95% CI [−987.00, −720.30]) and the Visual condition (798.45 ± 456.85 ms; $\beta = -187$, SE = 56.5, $t = -3.31$, 95% CI [−320.00, −53.70]). The reaction time for the Auditory condition was significantly longer than that for the Visual condition ($\beta = 667$, SE = 56.5, $t = 11.807$, 95% CI [533.00, 799.80]).

The main effect of Task type was also significant ($F(1, 220) = 22.85$, $p < .001$). Post-hoc comparisons showed that reaction time for the Repeating task was shorter than that for the Switching task ($\beta = -220$, SE = 46.1, $t = -4.78$, 95% CI [−311.00, −130.00]).

The current study did not find the significant main effect of language type, nor a significant interaction effect among the study variables (*ps* > 0.05) (Figure 2).

3.2.2. Event-related potentials

200–350 ms Window: The current study found a significant main effect of Modality ($F(2, 700) = 8.51$, $p < .001$). Post hoc comparisons showed that the Audiovisual condition (−0.01 ± 5.97 μV) elicited a smaller positive amplitude compared to the Auditory condition (0.87 ± 4.66 μV; $\beta = -0.89$, SE = 0.308, $t = -2.89$, 95% CI [−1.61, −0.17]) and the Visual condition (1.22 ± 4.41 μV; $\beta = -1.23$, SE = 0.308, $t = -4.00$, 95% CI [−1.96, −0.51]). There was no significant difference between the Auditory and Visual conditions ($p > .05$). We also found a significant main effect of Task type ($F(1, 700) = 8.01$, $p < .01$). Post hoc comparisons showed that the Repeating task elicited a smaller positive amplitude compared to the Switching task (Repeating task: 0.34 ± 5.44 μV, Switching task: 1.05 ± 6.26 μV; $\beta = -0.71$, SE = 0.252, $t = -2.83$, 95% CI [−1.21, −0.22]). The other effect was not significant.

350–700 ms Window: The current study found a significant main effect of Modality ($F(2, 700) = 13.02$, $p < .001$). Post hoc comparisons showed that the Audiovisual condition (0.52 ± 9.25 μV) elicited a smaller positive amplitude compared to the Visual condition (2.52 ± 6.73 μV; $\beta = -2.00$, SE = 0.418, $t = -4.78$, 95% CI [−2.98, −1.02]). The Auditory condition also elicited a smaller positive

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

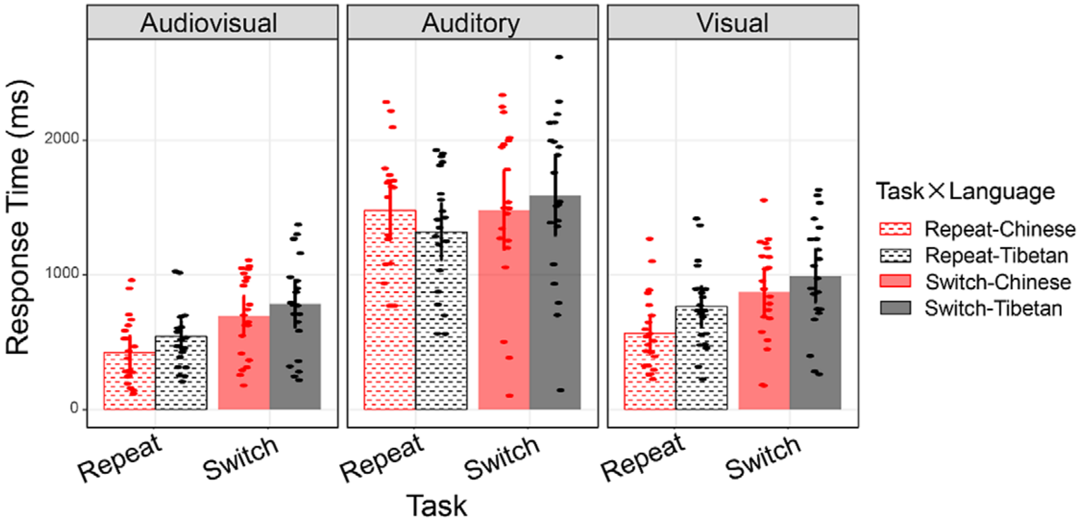

Figure 2. Response times (in milliseconds) for language switching and repetition tasks across three perceptual modalities (audiovisual, auditory and visual) in Tibetan–Chinese bilinguals. The bar plots represent mean response times, with error bars indicating standard error of the mean (SEM). The red bars correspond to tasks involving Chinese (L2), while the gray bars correspond to Tibetan (L1). The data points indicate individual participant responses.

amplitude compared to the Visual condition (Auditory: 0.88 ± 7.05 μV, Visual: 2.52 ± 6.72 μV; $\beta = -1.65$, SE = 0.418, $t = -3.94$, 95% CI [−2.63, −0.66]). The results further showed a significant interaction between ROI and Modality ($F(4, 700) = 8.39$, $p < .001$). Post hoc comparisons revealed that, in the medial anterior ROI, the Audiovisual condition (−1.99 ± 8.63 μV) elicited a smaller positive amplitude compared to the Auditory condition (0.59 ± 7.04 μV; $\beta = -2.58$, SE = 0.724, $t = -3.57$, 95% CI [−4.29, −0.88]) and the Visual condition (−0.18 ± 6.44 μV; $\beta = -1.81$, SE = 0.724, $t = -2.51$, 95% CI [−3.51, −0.11]). There was no significant difference in amplitude between the Auditory and Visual conditions ($p > .05$). In the medial central ROI, the Audiovisual condition (0.61 ± 7.92 μV) elicited a smaller positive amplitude compared to the Visual condition (2.72 ± 5.21 μV; $\beta = -2.11$, SE = 0.724, $t = -2.92$, 95% CI [−3.81, −0.41]). There was no significant difference in amplitude between the Auditory condition (1.44 ± 6.13 μV) and either the Visual (2.72 ± 5.21 μV) or Audiovisual (0.61 ± 7.92 μV) conditions (ps > 0.05). In the medial posterior ROI, both the Audiovisual condition (2.96 ± 6.56 μV; Audiovisual versus Visual: $\beta = -2.07$, SE = 0.724, $t = -2.86$, 95% CI [−3.77, −0.37]) and the Auditory condition (0.60 ± 4.99 μV; Auditory versus Visual: $\beta = -4.43$, SE = 0.724, $t = -6.12$, 95% CI [−6.13, −2.73]) elicited smaller positive amplitudes compared to the Visual condition (5.03 ± 4.30 μV). Additionally, we found that the Audiovisual condition elicited a larger positive amplitude compared to the Auditory condition ($\beta = 2.36$, SE = 0.724, $t = 3.27$, 95% CI [−6.13, −2.73]). We also found a significant interaction between Modality and Task type ($F(2, 700) = 3.29$, $p < .05$). Post hoc comparisons showed that, only in the Auditory modality, the Repeating task elicited a smaller positive amplitude compared to the Switching task (Repeating task: 0.16 ± 5.68 μV, Switching task: 1.60 ± 6.80 μV; $\beta = -1.45$, SE = 0.591, $t = -2.45$, 95% CI [−2.61, −0.29]) (Figure 3).

3.2.3. Results of neural prediction of switching cost

The best-performing model had an R-squared value of .35 and an MSE of 10,268.95. The 5000 permutation tests indicated that the best model performed significantly better than the R^2 generated with shuffled labels ($p = .0162$), and the MSE was significantly lower than that of the shuffled-label models ($p = .0002$). These results validate that the current model, based on neural features, predicted reaction times that matched well with the actual measured reaction times.

The top ten features contributing most to the model were: (1) Medial posterior region (200–350 ms), task: switching, modality: audiovisual, language: Tibetan (545.46); (2) Medial posterior region (200–350 ms), task: switching, modality: visual, language: Tibetan (519.51); (3) Medial anterior region (350–700 ms), task: switching, modality: audiovisual, language: Chinese (413.29); (4) Medial central region (200–350 ms), task: switching, modality: visual, language: Chinese (359.44); (5) Medial central region (200–350 ms), task: switching, modality: auditory, language: Chinese (327.75); (6) Medial anterior region (200–350 ms), task: switching, modality: audiovisual, language: Tibetan (314.45); (7) Medial posterior region (350–700 ms), task: switching, modality: visual, language: Tibetan (313.31); (8) Medial anterior region (350–700 ms), task: switching, modality: audiovisual, language: Tibetan (305.01); (9) Medial posterior region (200–350 ms), task: switching, modality: audiovisual, language: Chinese (304.57); and (10) Medial posterior region (200–350 ms), task: switching, modality: visual, language: Chinese (277.10) (Figures 4 and 5).

4. Study 2

Study 1 revealed main effects of Modality and Task type but failed to find any impact of Language type on switch costs. A close examination of its procedure showed that the language cues (a red frame for Tibetan versus a blue frame for Chinese) and their durations (500 ms versus 1000 ms) differed across Modality conditions. Such discrepancies may have affected cue processing and the switch costs associated with language control. Accordingly, in Study 2, we harmonized these discrepancies in cue type and duration to further explore the effects of language type and modality on switch costs.

4.1. Method

4.1.1. Participants

We recruited 20 university students proficient in both Tibetan and Mandarin (4 males; age = 20.40 ± 0.68 years). Participants self-reported high proficiency in both languages on a 7-point

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

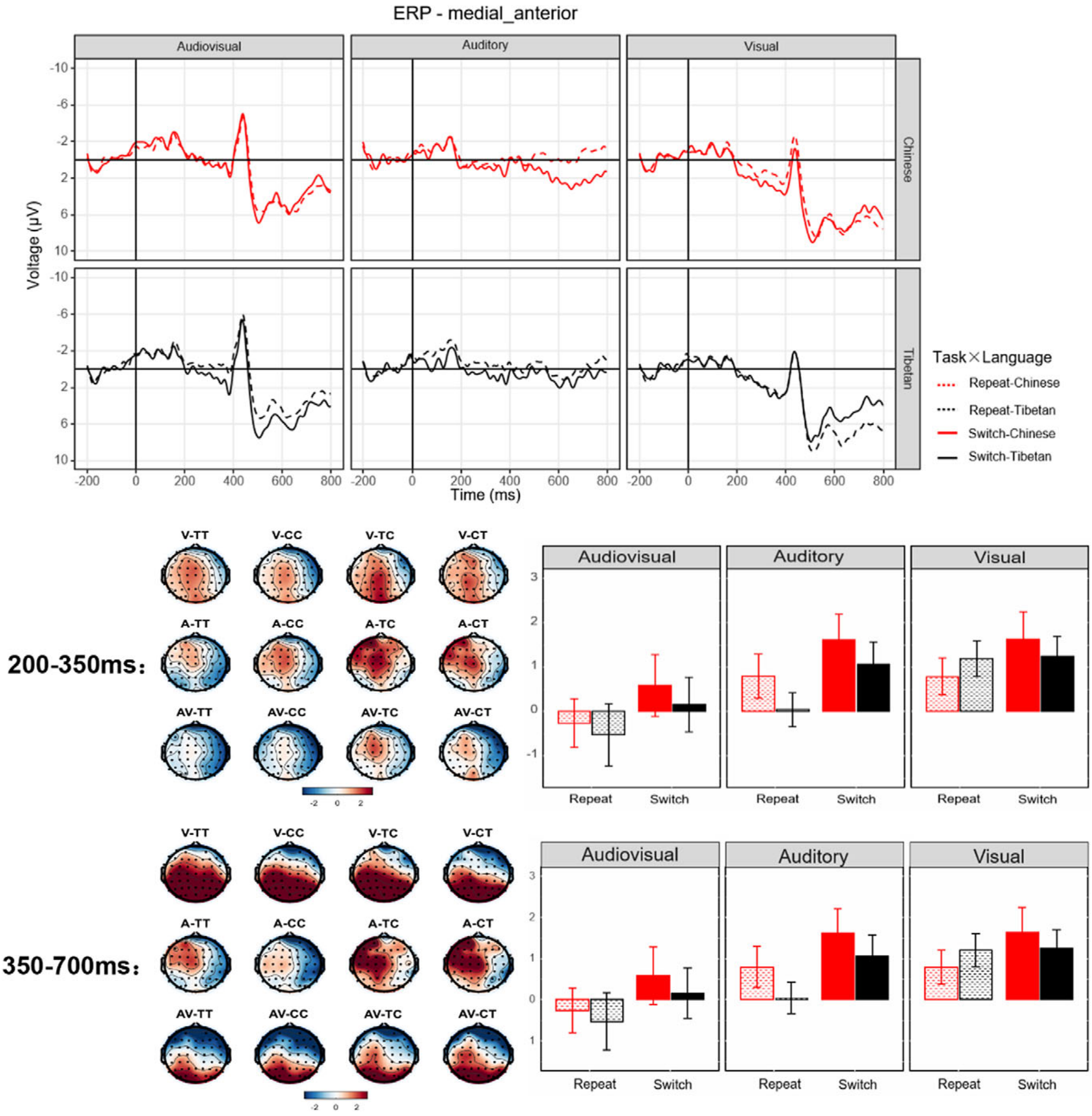

Figure 3. ERP Results for Bilingual Language Switching Tasks in Different Perceptual Modalities (Audiovisual, Auditory, Visual) in the Medial Anterior Region. The top panel displays ERP waveforms for repetition and switching tasks involving Tibetan (L1) and Chinese (L2). The red dashed line represents the Chinese repetition task, the gray dashed line represents the Tibetan repetition task, the solid red line represents the Chinese switching task and the solid black line represents the Tibetan switching task. The middle and bottom panels show topographic maps and bar plots of ERP responses within the early (200–350 ms) and late (350–700 ms) time windows across different task and language conditions (A = Auditory, V = Visual, AV = Audiovisual, CC = L1 Repetition Task, TT = L2 Repetition Task, CT = L1 to L2 Switching Task, TC = L2 to L1 Switching Task), with the region of interest (ROI) being the medial anterior area. The bar plots represent mean amplitudes, and error bars indicate the standard error of the mean (SEM).

Likert scale (1 = very low proficiency; 7 = very high proficiency), and there was no significant difference in proficiency between the two languages ($p > .05$). The present study used the same experimental design as Study 1, thereby achieving the ideal power level for the minimum required sample size.

4.1.2. Experimental materials and procedures

Experimental materials were identical to those in Study 1.

The procedures for the visual block and the audiovisual block were also the same as in Study 1.

In the auditory block, each trial began with a "+" presented for 500 ms, followed by a blank screen for 200 ms and then an audio clip lasting 1000 ms. Simultaneously with the audio clip, a red or blue frame appeared around the image for 1000 ms. A red frame cued participants to name the image in Tibetan (L1), whereas a blue frame cued naming in Mandarin (L2).

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

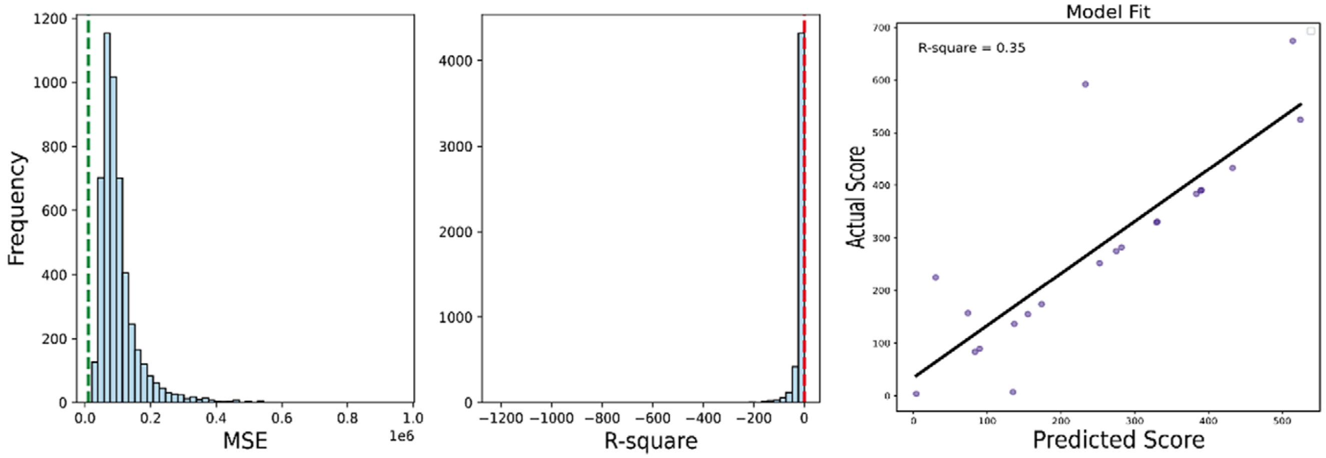

Figure 4. Machine learning model fitting results. The left panel shows the distribution of mean squared error (MSE) values across multiple iterations, with the green dashed line indicating the *p*-value result from the MSE permutation test. The middle panel illustrates the distribution of *R*-squared (R^2) values, with the red dashed line representing the *p*-value result from the R^2 permutation test. The right panel presents a scatter plot of the model's predicted scores versus actual scores, with an R^2 value of .35. The solid line represents the best-fit line.

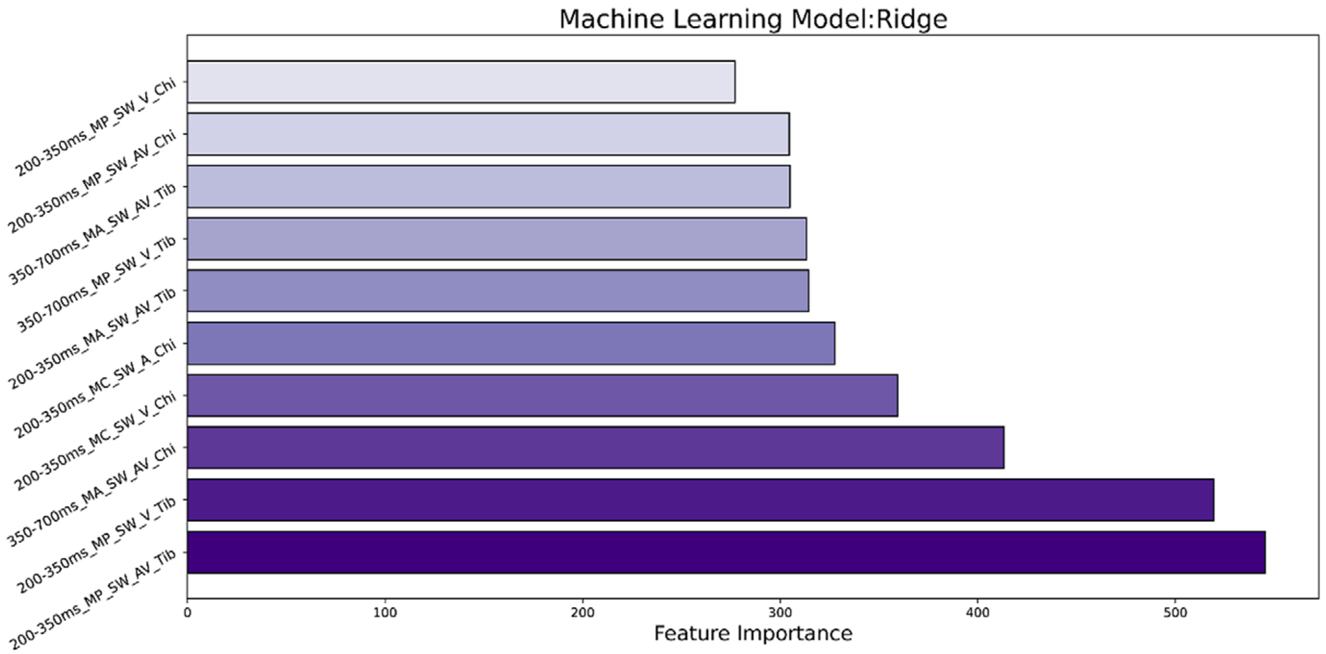

Figure 5. Machine learning feature importance results. The bar chart illustrates the feature importance values in the ridge machine learning model, indicating the relative contribution of different features in predicting language switching behavior. The darker bars represent higher feature importance. MP = medial posterior, MA = medial anterior, MC = medial central, SW = switching task, AV = audiovisual modality, V = visual modality, A = auditory modality, Tib = Tibetan, Chi = Chinese.

4.1.3. EEG recording and preprocessing

Same as in Study 1.

4.1.4. Data analysis

Behavioral data analysis followed the same procedures as in Study 1.

ERP data analysis followed the same procedures as in Study 1.

4.2. Results

4.2.1. Behavioral results

Modality showed a significant main effect, $F(2, 209) = 63.06$, $p < .001$. Post hoc comparisons revealed that reaction times were faster in the Audiovisual than in the Auditory condition (Audiovisual: 871.04 ± 403.61 ms; Auditory: 1107.73 ± 313.23 ms; $\beta = -237$, SE = 34, $t = -6.69$, 95% CI [−317, −156]), and that reaction times were slower in the Auditory than in the Visual condition (Visual: 729.85 ± 346.21 ms; Auditory: 1107.73 ± 313.23 ms; $\beta = 378$, SE = 34, $t = 11.11$, 95% CI [297.6, 458]). Task also showed a significant main effect, $F(1, 209) = 62.81$, $p < .001$. Reaction times on Switching task were significantly longer than on Repeating task (Switching task: 1012.90 ± 472.63 ms; Repeating task: 792.85 ± 403.70 ms; $\beta = 220$, SE = 27.8, $t = 7.93$, 95% CI [165, 275]).

4.2.2. Event-related potentials

200–350 ms Window: We observed a significant Language type × Modality interaction, $F(2, 665) = 4.42$, $p < .05$. Post hoc comparisons

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

showed that in the Tibetan condition, the Audiovisual stimulus elicited a smaller positive-going amplitude than the Auditory stimulus (Audiovisual: −0.12 ± 5.11 μV; Auditory: 0.81 ± 4.02 μV; $\beta = -0.93$, SE = 0.40, $t = -0.93$, 95% CI [−1.90, −0.04]), and that the Auditory stimulus elicited a larger positive-going amplitude than the visual stimulus (Visual: −0.50 ± 6.40 μV; Auditory: 0.81 ± 4.02 μV; $\beta = 1.31$, SE = 0.40, $t = 3.25$, 95% CI [0.343, 2.28]). Furthermore, the Modality × Language type × Task type interaction was significant, $F(4, 665) = 4.27$, $p < .05$. Post hoc tests revealed that only in the Auditory modality under the Chinese condition did repeat trials (versus switch trials) elicit a smaller positive-going amplitude (Repeating task: −0.67 ± 5.03 μV; Switching task: 0.92 ± 3.09 μV; $\beta = -1.59$, SE = 0.57, $t = -2.79$, 95% CI [−2.71, −0.47]).

350–700 ms Window: We found a significant ROI × Modality interaction, $F(4, 700) = 8.93$, $p < .001$. Post hoc comparisons showed that in the medial anterior ROI, Audiovisual stimuli elicited a smaller positive-going amplitude than Auditory stimuli (Audiovisual: −2.32 ± 7.37 μV; Auditory: 0.71 ± 6.57 μV; $\beta = -3.03$, SE = 0.69, $t = -4.38$, 95% CI [−4.69, −1.37]), and that Auditory stimuli elicited a larger positive-going amplitude than Visual stimuli (Visual: −1.91 ± 7.86 μV; Auditory: 0.71 ± 6.57 μV; $\beta = 2.61$, SE = 0.69, $t = 3.78$, 95% CI [0.95, 4.27]). In the medial posterior ROI, Audiovisual stimuli elicited a larger positive-going amplitude than Auditory stimuli (Audiovisual: 2.19 ± 5.04 μV; Auditory: 0.22 ± 4.58 μV; $\beta = 1.97$, SE = 0.69, $t = 2.85$, 95% CI [0.31, 3.63]), and Auditory stimuli elicited a smaller positive-going amplitude than Visual stimuli (Visual: 2.40 ± 4.36 μV; Auditory: 0.22 ± 4.58 μV; $\beta = -2.18$, SE = 0.69, $t = -3.15$, 95% CI [−3.84, −0.52]). We also observed a significant Language type × Modality interaction, $F(2, 700) = 5.90$, $p < .01$. Post hoc tests revealed that in the Tibetan condition, Audiovisual stimuli elicited a smaller positive-going amplitude than Auditory stimuli (Audiovisual: −0.63 ± 7.19 μV; Auditory: 1.10 ± 5.67 μV; $\beta = -1.74$, SE = 0.56, $t = -3.08$, 95% CI [−3.09, −0.38]), and that Auditory stimuli elicited a larger positive-going amplitude than Visual stimuli (Visual: −0.53 ± 7.52 μV; Auditory: 1.10 ± 5.67 μV; $\beta = 1.64$, SE = 0.56, $t = 2.90$, 95% CI [0.28, 2.99]). Furthermore, the Modality × Language type × Task type interaction was marginally significant, $F(2, 700) = 2.76$, $p = .06$. Post hoc comparisons showed that only in the auditory modality under the Chinese condition did Repeating task (versus Switching task) elicit a smaller positive-going amplitude (Switching task: −1.16 ± 7.18 μV; Repeating task: 0.81 ± 3.78 μV; $\beta = -1.97$, SE = 0.80, $t = -2.47$, 95% CI [−3.54, −0.40]).

5. Discussion

Study 1 investigates the language switching process of Tibetan–Chinese bilinguals in different perceptual modalities (visual, auditory and audiovisual), analyzing behavioral responses and electrophysiological signals (ERPs). The results showed that, in terms of behavioral performance, the reaction time for language switching was fastest in the audiovisual modality stimulus and slowest in the auditory modality stimulus, indicating that the audiovisual modality stimulus may facilitate faster language processing and switching. Additionally, the reaction time for repeated tasks was significantly shorter than for switching tasks, consistent with previous research, suggesting that task continuity helps reduce the difficulty of language switching (Declerck et al., 2021; Gullifer & Titone, 2019; Lavric et al., 2019). In terms of ERPs, the positive amplitude elicited by the audiovisual modality stimulus was significantly lower than that elicited by the auditory and visual modalities, as observed in both the 200–350 ms and 350–700 ms time windows. This indicates that the audiovisual modality stimulus requires fewer neural resources during language processing, potentially making the brain more efficient in managing language switching. Furthermore, multimodal sensory information may have a positive impact on language control. Machine learning analysis results showed that when stimuli were presented in the audiovisual modality, with the language switching direction being from Chinese to Tibetan, and when early ERP signals were recorded in the posterior region of interest (ROI), these neural signal features were most important for predicting the switching cost of the second language (i.e., behavioral performance in switching from Chinese to Tibetan). Study 2, by standardizing the cue signals across different perceptual modalities, replicated the main effects of Task type and Modality on both behavioral and ERP outcomes observed in Study 1; on the other hand, unlike Study 1, significant switch costs only emerged in the auditory modality when switching to the second language in the 200–350 ms and 350–700 ms windows.

5.1. Effects of modality type on language switching: evidence from behavioral and ERP data

The behavioral results of this study demonstrated that reaction times increased sequentially in audiovisual, visual and auditory conditions, indicating significant differences in processing speed and task difficulty across different perceptual modalities. In the early ERP time window, neural activity showed a pattern of separation, with evoked amplitudes becoming increasingly positive across audiovisual, visual and auditory conditions. In the current study, we observed an N2-like effect in the 200–350 ms time window that was consistent with Lavric et al. (2019), indicating that the switch condition elicited a stronger N2-like effect than the repeat condition. We further found that the audiovisual modality stimulus provided more information in language production tasks than the unimodal visual or auditory modalities. This is because the audiovisual modality integrates multiple visual and auditory inputs to more effectively support language generation. During language switching, we propose that audiovisual information helps reduce the switching cost, thereby leading to a smaller N2-like effect. There are two reasons for this: First, in a previous verbal Stroop task with four conditions (congruent, pure auditory and pure visual), the task examined the impact of sensory integration on cognitive conflict by presenting different visual and auditory stimuli, and the results showed that response times in the audiovisual condition were faster than those in the pure auditory and pure visual conditions (Fitzhugh et al., 2019); Second, we conducted an exploratory analysis using the neural signal of the N2-like effect in the switch condition of the audiovisual modality as a predictor to jointly predict the switching cost in the audiovisual modality. The results indicated that, in the Chinese switch task, the audiovisual N2-like effect in the posterior region significantly predicted the switching cost in the audiovisual modality ($\beta = 127.06$, $t = 2.279$, $p = .0389$, as shown in Figure S1). This predictive effect was not observed in the visual and auditory modalities.

In the subsequent late time window (350–700 ms), we observed significant differences in how different perceptual modalities modulated task types. Specifically, only in the auditory modality stimulus did the language switching task evoke more positive amplitudes compared to the repetition task. This result suggests that individuals require more neural resources to achieve language control when switching languages in auditory conditions. Consistent with previous

https://doi.org/10.1017/S1366728925100424 Published online by Cambridge University Press

research, the amplitude of the late positive component (LPC) is closely associated with greater demands for language control (Lavric et al., 2019). In the current study, we found that language switching tasks in auditory conditions required more language control compared to repetition tasks, indicating that different perceptual modalities may lead to asymmetrical switch costs. This asymmetry manifested in the neural signals observed in this study, where significant differences between switching and repetition tasks were found only in the auditory condition, whereas no such differences were observed in visual or audiovisual conditions. According to the ICM, language switching in auditory conditions involves greater neural control demands (Declerck et al., 2015, 2021; D. W. Green, 1998). Auditory language processing tends to be more complex than visual processing because auditory signals need to be integrated over time and lack spatial cues that can aid comprehension, thus increasing the demand for language control. This leads to higher switch costs and more positive amplitudes. On the other hand, in the audiovisual condition, the dual input of both visual and auditory information allows individuals to alleviate the need for inhibition of the nontarget language by integrating multimodal information, thereby reducing the difficulty of language switching. Consequently, language switching costs were lower and reaction times were shorter in audiovisual conditions. This finding suggests that multimodal perceptual information may play a facilitating role in language processing, alleviating the burden of inhibitory control. Moreover, by standardizing both the cue mode and its timing across modalities in Study 2, the results further showed that processing the second language in the auditory modality induces larger switch costs. The design of Study 2 parallels that of Xing Qiang (2021), in which the target-language sequence was unpredictable; thus, participants could not know in advance which language would be required on each trial and could not preactivate the corresponding language system in the brain. The current findings suggest that low-level processes (such as decoding auditory input and mapping it to phonological representations) occur concurrently with the inhibition–activation mechanisms needed for language switching, resulting in stronger interference (Hayes-Harb et al., 2010).

5.2. Neurodynamic features of Tibetan–Chinese bilingual language switching

In this study, we found no significant interaction between language type (Tibetan versus Chinese) and task type (repetition versus switching) in either behavioral or neural indicators. This indicates that there is no significant difference in bilingual switch costs when switching from Tibetan to Chinese compared to switching from Chinese to Tibetan. This result differs from the asymmetrical switch costs observed in previous studies. A possible reason for this is that the participants in this study were Tibetan–Chinese university students who were highly proficient in both languages. They reported no significant differences in their proficiency levels for Tibetan (L1) and Chinese (L2) across reading, writing, speaking and comprehension skills.

Previous research has shown that highly proficient bilinguals exhibit symmetric switch costs because the activation levels of their two languages are similar, which reduces reliance on inhibitory mechanisms (Costa & Santesteban, 2004; Schwieter & Sunderman, 2008). For example, Meuter and Allport (1999) found that high-proficiency bilinguals showed symmetric costs in language switching, whereas low-proficiency bilinguals exhibited asymmetrical switch costs. Additionally, Filippi et al. (2014) found that asymmetrical switch costs were negatively correlated with L2 proficiency, indicating that as proficiency in the second language increased, switch costs tended to become symmetric. These findings further support our results showing symmetric switch costs in Tibetan–Chinese bilinguals. According to the ICM, language switching costs primarily arise from sustained inhibition mechanisms (Green, 1998). When one trial involves using a particular language, the nontarget language is inhibited. In the subsequent trial, if the previously inhibited language is required, the lingering inhibition must be overcome, which increases the language switching load and makes language switching more difficult. In contrast, when repeating the same language, the target language remains activated and the nontarget language continues to be inhibited, resulting in faster reaction times and higher accuracy. ICM also posits that the inhibition process is reactive, i.e., the higher the activation level, the stronger the inhibition required. Therefore, when switching to a previously highly activated and strongly inhibited language, the difficulty of releasing inhibition increases, leading to higher switch costs. However, for individuals with high L2 proficiency, the activation levels of both languages are more balanced, which likely reduces the need for inhibition, resulting in symmetric switch costs (Declerck & Philipp, 2015). This suggests that language proficiency significantly affects the efficiency of inhibitory mechanisms and the fluency of language switching.

5.3. Neural signals predicting L1–L2 switch costs: results based on machine learning

The current study found that audiovisual modality, L2–L1 language switching direction and early time-window neural signals in the posterior region of interest played a crucial role in predicting L1–L2 switch costs. This suggests that the mode of stimulus presentation, the direction of language switching and early neural signals are all vital for predicting switch costs in language processing.

Previous research has separately compared the effects of visual and auditory modalities on language switching, revealing distinct switching behaviors in bilinguals in different conditions: in some cases, the cost of visual switching was greater than auditory, while in others, the opposite was true (Declerck et al., 2015; Xing Qiang, 2021). However, few studies have explored the impact of audiovisual modality stimulus (i.e., the combination of visual and auditory) on bilingual switching. In the present study, behavioral results indicated that compared to the single visual or auditory modality stimulus, the audiovisual modality stimulus significantly accelerated overall language processing speed, both in repetition and language switching tasks. This suggests that audiovisual stimuli might enhance language processing efficiency through its multimodal information, thereby promoting smoother bilingual switching. Some reviews have highlighted that audiovisual stimuli can induce gamma oscillations and have potential benefits in treating cognitive, emotional and sleep disorders (Chen et al., 2022). Furthermore, audiovisual content in daily life, such as television and films, plays an important role in language acquisition, particularly when learners are exposed without subtitles or with different types of subtitles (interlingual and intralingual), providing an effective language environment for learners (Caruana, 2021; Montero Perez, 2022; Zhang & Zou, 2022). An empirical study also indicated that people find audiovisual modality stimulus more attractive than single visual or auditory input (Hu et al., 2023). These findings suggest that the rich information available in audiovisual modality stimulus is particularly valuable for second language learners. Moreover, the study found that neural activity evoked by L2–L1 language switching could significantly predict L1–L2 switching behavior, implying that similar neural signals may be present for both switching directions during the early time window. This finding is consistent with our experimental results, which showed no significant asymmetry between the two switching directions in terms of switch costs at the behavioral and neural levels (Declerck et al., 2021). Most importantly, the study emphasized the importance of early time-window neural signals for predicting language switching behavior. This conclusion aligns with previous research on the neurodynamics of bilingual switching, which found a main effect of switch cost (switch versus repeat) at around 200–300 ms (Peeters, 2020) and 200–350 ms (Declerck et al., 2021; Verhoef et al., 2010). These results further support the crucial role of early neural responses in bilingual language switching.

We additionally constructed four ridge regression models, namely: Model 1: Using neural signals from the audiovisual modality to predict switching behavior in the audiovisual modality (see Table S1 in the Supplementary Materials); Model 2: Using neural signals from the visual modality to predict switching behavior in the visual modality (see Table S2 in the Supplementary Materials); Model 3: Using neural signals from the auditory modality to predict switching behavior in the auditory modality (see Table S4 in the Supplementary Materials) and Model 4: Using unimodal neural signals from the visual and auditory modalities to predict switching behavior in the audiovisual modality (see Table S4 in the Supplementary Materials). Among the within-modality predictive models (i.e., Models 1, 2 and 3), the early time window plays a particularly important role in predicting the corresponding labels for models that include visual signals, whereas for the model that relies solely on auditory signals, the late time window is critical. In addition, the ROIs in the within-modality predictive models are primarily concentrated in the central and frontal areas. Regarding the cross-modality predictive model (i.e., Model 4), visual (versus auditory) signals are most critical for predicting the switching cost of multimodal cues, with the early time window in the central and posterior regions also playing an important role. This finding is consistent with previous studies suggesting that visual signals are more important than auditory signals in the perception of multimodal social cues (Hu et al., 2023).

6. Conclusion

Study 1 utilized a cue-based bilingual switching task to explore the impact of audiovisual modality stimulus on the language switching behavior and neurodynamics of Tibetan–Chinese bilinguals. Behavioral results demonstrated that, compared to the single visual or auditory modalities, the audiovisual modality significantly enhanced language output, highlighting the importance of multisensory input in language processing. ERP results further revealed that audiovisual modality evoked a smaller positive amplitude in the early time window (200–350 ms), indicating reduced neural activity compared to visual and auditory conditions. Additionally, in the later time window (350–700 ms), only in the auditory modality did the neural signals for repetition and switching tasks exhibit a significant divergence, reflecting the unique neural processing characteristics of the auditory modality stimulus. Machine learning analysis showed that audiovisual modality, L2–L1 language switching direction and early time-window neural signals in the posterior region of interest were most important in predicting L1–L2 language switching behavior. In the auditory modality, we observed an asymmetrical switch cost, suggesting that auditory input may influence the balance of language switching in specific contexts. These results were replicated in Study 2, and it was further found that processing the second language in the auditory modality affects switch costs. The findings demonstrate that the audiovisual modality reduces language switch costs, whereas the auditory modality and second-language processing increase them.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100424.

Data availability statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding statement. This work was supported by the National Social Science Fund of China (23XMZ050) and the scientific and technological innovation 2030 – the major project of the Brain Science and Brain-Inspired Intelligence Technology (2021ZD0200500). Additional funding was provided by the Graduate Program of Northwest Normal University (2024–2025).

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Asaadi, A. H., **Amiri, S. H.**, **Bosaghzadeh, A.**, & **Ebrahimpour, R.** (2024). Effects and prediction of cognitive load on encoding model of brain response to auditory and linguistic stimuli in educational multimedia. *Scientific Reports*, **14**(1), 9133. https://doi.org/10.1038/s41598-024-59411-x.

Athanasopoulos, P., **Bylund, E.**, **Montero-Melis, G.**, **Damjanovic, L.**, **Schartner, A.**, **Kibbe, A.**, **Riches, N.**, & **Thierry, G.** (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, **26**(4), 518–526. https://doi.org/10.1177/0956797614567509.

Calabria, M., **Jacquin-Courtois, S.**, **Miozzo, A.**, **Rossetti, Y.**, **Padovani, A.**, **Cotelli, M.**, & **Miniussi, C.** (2011). Time perception in spatial neglect: A distorted representation? *Neuropsychology*, **25**(2), 193. https://doi.org/10.1037/a0021304.

Caruana, S. (2021). An overview of audiovisual input as a means for foreign language acquisition in different contexts. *Language and Speech*, **64**(4), 1018–1036. https://doi.org/10.1177/0023830920985897.

Çekiç, A. (2024). Incidental L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*, **37** (4), 896–923. https://doi.org/10.1080/09588221.2022.2062004.

Chen, X., **Shi, X.**, **Wu, Y.**, **Zhou, Z.**, **Chen, S.**, **Han, Y.**, & **Shan, C.** (2022). Gamma oscillations and application of 40-Hz audiovisual stimulation to improve brain function. *Brain and Behavior*, **12**(12), e2811. https://doi.org/10.1002/brb3.2811.

Christoffels, I. K., **Firk, C.**, & **Schiller, N. O.** (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, **1147**, 192–208. https://doi.org/10.1016/j.brainres.2007.01.137.

Costa, A., & **Santesteban, M.** (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, **50**(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002.

Costa, A., **Santesteban, M.**, & **Ivanova, I.** (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **32**(5), 1057. https://doi.org/10.1037/0278-7393.32.5.1057.

Declerck, M., **Meade, G.**, **Midgley, K. J.**, **Holcomb, P. J.**, **Roelofs, A.**, & **Emmorey, K.** (2021). On the connection between language control and executive control—An ERP study. *Neurobiology of Language*, **2**(4), 628–646. https://doi.org/10.1162/nol_a_00032.

Declerck, M., & **Philipp, A. M.** (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, **22**(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1.

Declerck, M., **Stephan, D. N.**, **Koch, I.**, & **Philipp, A. M.** (2015). The other modality: Auditory stimuli in language switching. *Journal of Cognitive Psychology*, **27**(6), 685–691. https://doi.org/10.1080/20445911.2015.1026265.

Faul, F., **Erdfelder, E.**, **Lang, A.-G.**, & **Buchner, A.** (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**(2), 175–191. https://doi.org/10.3758/BF03193146.

Filippi, R., **Karaminis, T.**, & **Thomas, M. S. C.** (2014). Language switching in bilingual production: Empirical data and computational modelling. *Bilingualism: Language and Cognition*, **17**(2), 294–315. https://doi.org/10.1017/S1366728913000485.

Fitzhugh, M. C., **Whitehead, P. S.**, **Johnson, L.**, **Cai, J. M.**, **Baxter, L. C.**, & **Rogalsky, C.** (2019). A functional MRI investigation of crossmodal interference in an audiovisual stroop task. *PLoS One*, **14**(1), e0210736. https://doi.org/10.1371/journal.pone.0210736.

Flege, J. E., & **Bohn, O.-S.** (2021). The revised speech learning model (SLM-r). *Second Language Speech Learning: Theoretical and Empirical Progress*, **10** (9781108886901.002).

Gao, Y., & **Zeng, G.** (2021). An exploratory study on national language policy and family language planning in the Chinese context. *Cogent Education*, **8**(1), 1878871. https://doi.org/10.1080/2331186X.2021.1878871.

Goldrick, M., & **Gollan, T. H.** (2023). Inhibitory control of the dominant language: Reversed language dominance is the tip of the iceberg. *Journal of Memory and Language*, **130**, 104410. https://doi.org/10.1016/j.jml.2023.104410.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**(2), 67–81. https://doi.org/10.1017/S1366728998000133.

Green, P., & **MacLeod, C. J.** (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, **7**(4), 493–498. https://doi.org/10.1111/2041-210X.12504.

Gullifer, J. W., & **Titone, D.** (2019). The impact of a momentary language switch on bilingual reading: Intense at the switch but merciful downstream for L2 but not L1 readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **45**(11), 2036–2050. https://doi.org/10.1037/xlm0000695.

Gullifer, J. W., & **Titone, D.** (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, **23**(2), 283–294. https://doi.org/10.1017/S1366728919000026.

Hayes-Harb, R., **Nicol, J.**, & **Barker, J.** (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, **53**(3), 367–381. https://doi.org/10.1177/0023830910371460.

Hu, Y., **Li, R.**, **Jiang, X.**, & **Chen, W.** (2024). The change in aesthetic experience and empathic concern predicts theory of mind ability: Evidence from drama improvisation training. *The Arts in Psychotherapy*, **89**, 102167. https://doi.org/10.1016/j.aip.2024.102167.

Hu, Y., **Mou, Z.**, & **Jiang, X.** (2023). Attentional relevance modulates nonverbal attractiveness perception in multimodal display. *Journal of Nonverbal Behavior*, **47**(3), 285–319. https://doi.org/10.1007/s10919-023-00428-7.

Kheder, S., & **Kaan, E.** (2021). Cognitive control in bilinguals: Proficiency and code-switching both matter. *Cognition*, **209**, 104575. https://doi.org/10.1016/j.cognition.2020.104575.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford University Press.

Lavric, A., **Clapp, A.**, **East, A.**, **Elchlepp, H.**, & **Monsell, S.** (2019). Is preparing for a language switch like preparing for a task switch? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **45**(7), 1224. https://doi.org/10.1037/xlm0000636.

Lavric, A., **Mizon, G. A.**, & **Monsell, S.** (2008). Neurophysiological signature of effective anticipatory task-set control: A task-switching investigation. *European Journal of Neuroscience*, **28**(5), 1016–1029. https://doi.org/10.1111/j.1460-9568.2008.06372.x.

Linck, J. A., **Schwieter, J. W.**, & **Sunderman, G.** (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, **15**(3), 651–662. https://doi.org/10.1017/S136672891100054X.

Liu, S., **Huang, J.**, **Xing, Z.**, **Schwieter, J. W.**, & **Liu, H.** (2024). *Neural correlates of compound head position in language control: Evidence from simultaneous production and comprehension* (pp. 1–13). Bilingualism: Language and Cognition. https://doi.org/10.1017/S1366728923000883.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, **49**(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y.

Meuter, R. F. I., & **Allport, A.** (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, **40**(1), 25–40. https://doi.org/10.1006/jmla.1998.2602.

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, **55**(2), 163–192. https://doi.org/10.1017/S0261444821000501.

Morett, L. M., **Feiler, J. B.**, & **Getz, L. M.** (2022). Elucidating the influences of embodiment and conceptual metaphor on lexical and non-speech tone learning. *Cognition*, **222**, 105014. https://doi.org/10.1016/j.cognition.2022.105014.

Muñoz, C., **Pujadas, G.**, & **Pattemore, A.** (2023). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, **39**(1), 13–37. https://doi.org/10.1177/02676583211015797.

Ou, J., & **Law, S.-P.** (2017). Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception, & Psychophysics*, **79**(3), 945–963. https://doi.org/10.3758/s13414-017-1283-z.

Peeters, D. (2020). Bilingual switching between languages and listeners: Insights from immersive virtual reality. *Cognition*, **195**, 104107. https://doi.org/10.1016/j.cognition.2019.104107.

Pereira Soares, S. M., **Prystauka, Y.**, **DeLuca, V.**, **Poch, C.**, & **Rothman, J.** (2024). Brain correlates of attentional load processing reflect degree of bilingual engagement: Evidence from EEG. *NeuroImage*, **298**, 120786. https://doi.org/10.1016/j.neuroimage.2024.120786.

Rajan, M. P. (2022). An efficient ridge regression algorithm with parameter estimation for data analysis in machine learning. *SN Computer Science*, **3**(2), 171. https://doi.org/10.1007/s42979-022-01051-x.

Schwieter, J. W., & **Sunderman, G.** (2008). Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon*, **3**(2), 214–238. https://doi.org/10.1075/ml.3.2.06sch.

Seitz, S. R., & **Smith, S. A.** (2022). Talking the talk: Considering forced language-switching in the workplace. *Human Resource Management Review*, **32**(2), 100833. https://doi.org/10.1016/j.hrmr.2021.100833.

Smith, S. A., **Seitz, S. R.**, **Koutnik, K. H.**, **McKenna, M.**, & **Garcia, J. N.** (2020). The “work” of being a bilingual: Exploring effects of forced language switching on language production and stress level in a real-world setting. *Applied PsychoLinguistics*, **41**(3), 701–725. https://doi.org/10.1017/S0142716420000259.

Venkatesh, K. A., **Mishra, D.**, & **Manimozhi, T.** (2023). 9—Model selection and regularization. In T. Goswami & G. R. Sinha (Eds.), *Statistical Modeling in machine learning* (pp. 159–178). Academic Press.

Verhoef, K. M. W., **Roelofs, A.**, & **Chwilla, D. J.** (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, **22**(8), 1832–1843. https://doi.org/10.1162/jocn.2009.21291.

Wang, Y., & **Wei, L.** (2023). Multilingual learning and cognitive restructuring: The role of audiovisual media exposure in Cantonese–English–Japanese multilinguals’ motion event cognition. *International Journal of Bilingualism*, **27**(3), 331–348. https://doi.org/10.1177/13670069221085565.

Xing Qiang, W. U. X. W. J. Z. Z. (2021). The influence of the matching of modality presentation mode and perceptual learning style on the bidialectal switching cost of Cantonese-mandarin. *Acta Psychologica Sinica*, **53**(10), 1059–1070. https://doi.org/10.3724/SP.J.1041.2021.01059.

Zhang, R., & **Zou, D.** (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, **35**(9), 2790–2816. https://doi.org/10.1080/09588221.2021.1896555.