Brain functional plasticity associated with the emergence of expertise in extreme language control

CrossMark

Alexis Hervais-Adelman [a,*], Barbara Moser-Mercer [b], Narly Golestani [a]

[a] Brain and Language Lab, Department of Clinical Neuroscience, University of Geneva, Switzerland
[b] Faculty of Translation and Interpreting, University of Geneva, Switzerland

ABSTRACT

We used functional magnetic resonance imaging (fMRI) to longitudinally examine brain plasticity arising from long-term, intensive simultaneous interpretation training. Simultaneous interpretation is a bilingual task with heavy executive control demands. We compared brain responses observed during simultaneous interpretation with those observed during simultaneous speech repetition (shadowing) in a group of trainee simultaneous interpreters, at the beginning and at the end of their professional training program. Age, sex and language-proficiency matched controls were scanned at similar intervals.

Using multivariate pattern classification, we found distributed patterns of changes in functional responses from the first to second scan that distinguished the interpreters from the controls. We also found reduced recruitment of the right caudate nucleus during simultaneous interpretation as a result of training. Such practice-related change is consistent with decreased demands on multilingual language control as the task becomes more automatized with practice. These results demonstrate the impact of simultaneous interpretation training on the brain functional response in a cerebral structure that is not specifically linguistic, but that is known to be involved in learning, in motor control, and in a variety of domain-general executive functions. Along with results of recent studies showing functional and structural adaptations in the caudate nuclei of experts in a broad range of domains, our results underline the importance of this structure as a central node in expertise-related networks.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Simultaneous interpretation is a challenging task that places extreme demands on the cognitive control of language, and on a variety of executive functions including verbal working memory and attention. It requires simultaneous encoding of an incoming speech stream in one language, conversion of the content of that stream to a lexically, semantically and syntactically valid form in the target language, production in the target language and ongoing monitoring of the output to ensure message equivalence between source and target languages (Moser, 1978; Moser-Mercer et al., 1997).

Learning to become a simultaneous interpreter does not depend solely on the acquisition of linguistic expertise; trainee interpreters already master the languages in which they intend to work. Rather, their training involves the acquisition of a high degree of control over their language management skills, as well as the development of a non-linguistic cognitive toolkit that ensures mastery of divided attention (attending to speech input and output simultaneously) and efficient use of working memory, in order to enable them to rapidly and dynamically produce an accurate reproduction of the source message in diverse contexts. We consider that simultaneous interpretation requires the application of mechanisms of polyglot language control, and that training in SI is likely to train these mechanisms. There is ongoing debate about whether language selection takes place via a process of inhibition (e.g. Abutalebi and Green, 2007; Green, 1998; Prior; Van Heuven et al., 2008), by the maintenance of goals and self-monitoring (Colzato et al., 2008; Costa et al., 2009) or a combination of these (Bialystok et al., 2012; Green and Abutalebi, 2013). Similarly, the neural bases of bilingual language control are not yet fully elucidated. Structural brain analyses have shown differences between the brains of bilingual and monolingual individuals in the inferior parietal lobe (Grogan et al., 2012; Mechelli et al., 2004), in the left putamen (Abutalebi et al., 2013), in the anterior cingulate cortex (Abutalebi et al., 2012), in the left caudate nucleus of bimodal bilinguals (Zou et al., 2012) and in the left inferior frontal gyrus (Klein et al., 2014; Stein et al., 2012). Functional imaging work has suggested a role for a number of language-related and cognitive control-related regions including the striatum (e.g. Crinion et al., 2006; Klein et al., 1994, 1995, 2006), the dorsolateral prefrontal cortices, the inferior frontal gyrus, and parietal regions, as reviewed extensively by a number of authors (see, for example Costa

---

and Sebastian-Galles, 2014; Friederici, 2006; Hervais-Adelman et al., 2011; and a meta-analysis by Luk et al., 2012). The apparent relationship between mechanisms of language control and cognitive control is increasingly suggested to be at the root of the development of superior executive functions in bilinguals (Costa and Sebastian-Galles, 2014; Stocco et al., 2012).

Behavioral investigations of the cognitive benefits of expertise in simultaneous interpretation have revealed cognitive advantages beyond those thought to be attributable to bilingualism per se. Evidence points towards advantages on working memory (Christoffels et al., 2006; Daró and Fabbro, 1994; Fabbro and Daró, 1995; Kopke and Nespoulous, 2006; Stavrakaki et al., 2012; Yudes et al., 2011, 2012), and on some aspects of cognitive control, such as cognitive flexibility (Yudes et al., 2011). Little work exists on linguistic-processing differences between interpreters and non-interpreter multilinguals, although one EEG study has reported different cross-linguistic semantic priming effects in professional interpreters comparing to proficiency-matched L2 speakers (Elmer et al., 2010).

A recent fMRI study by Hervais-Adelman and colleagues (Hervais-Adelman et al., 2014) investigated the network of brain areas implicated in simultaneous interpretation in naïve participants. The study revealed a broad network of regions, including the pre-supplementary motor area, the dorsal anterior cingulate, the left inferior frontal gyrus, the left posterior temporal areas, the left anterior insula, the left superior and inferior cerebellar lobules as well as the caudate nuclei and the putamen. The involvement of the basal ganglia in this task is particularly interesting, not only in that it is consistent with an increasing body of evidence from neuroimaging, neuropsychology and electrical stimulation studies that have shown involvement of these structures in multilingual language control (Friederici, 2006; Hervais-Adelman et al., 2011), but also because of their involvement in domain-general cognitive control (Aron et al., 2007a,b; Grahn et al., 2008, 2009; Graybiel, 1995a,b; Houk et al., 2007; Moritz-Gasser and Duffau, 2009). Hervais-Adelman and colleagues suggested that the caudate nuclei respond to the overall demands of maintaining the lexico-semantic sets of two languages 'live' simultaneously during interpretation, while the right putamen is responsible for the moment-to-moment control of language, ensuring speech production in the appropriate language. These results contrast with those of an earlier functional neuroimaging, PET study which was carried out in 8 expert simultaneous interpreters having had between 5 and 20 years of professional experience. Here, recruitment of left premotor and dorsolateral prefrontal cortices during simultaneous interpretation was found. The fewer regions revealed in this study compared to ours can be due to the small sample size in this former study, and to the fact that in this former study, professionals were tested whereas in ours, naïve participants were tested. Indeed, fMRI research on the impact of practice suggests that reductions in recruitment of brain areas during task performance can occur as a result of training (Kelly and Garavan, 2005). Furthermore, work on expertise has also suggested that reductions in brain activity might be seen in expert compared to novice performance. For example, studies on expert performance in professional golfers planning shots (Milton et al., 2007) and bimanual co-ordination in professional concert pianists (Haslinger et al., 2004) both showed reduced brain activation in experts compared to controls.

Two studies have examined the structural bases of expertise in simultaneous interpretation. One diffusion tensor imaging (DTI) study by Elmer and colleagues has shown white matter differences between 12 interpreters and 12 non-expert control participants in brain regions associated with sensorimotor coupling and interhemispheric information transfer, including portions of the anterior and posterior corpus callosum, the left insula, and the right caudate nucleus (Elmer et al., 2011). In a further study by the same group, seemingly carried out on the same participants as the above-described DTI study, reduced gray matter volumes were found in professional simultaneous interpreters compared to multilingual controls in the left middle-anterior cingulate

gyrus, bilateral pars triangularis, left pars opercularis, bilateral middle part of the insula, and in the left supramarginal gyrus. Further, within the group of simultaneous interpreters, negative correlations were found between cumulative hours of practice of SI and gray-matter volume in the left pars triangularis, left middle anterior cingulate, and bilaterally in the caudate nucleus (Elmer et al., 2014).

In this study, we used fMRI to investigate brain functional plasticity in trainee simultaneous interpreters learning to become professional conference interpreters. We examined cerebral responses during simultaneous interpretation before and after a fifteen-month period of intensive training in a group of 19 trainee interpreters. We compared these to the cerebral responses of 16 multilingual control participants, tested on the same tasks over the same interval. The control participants were students engaged in a course of higher-level study in non-linguistic domains. Our study is the first longitudinal brain imaging study in simultaneous interpreters, and as such, will provide more direct evidence for training-induced brain plasticity than previous cross-sectional investigations. We expect that the acquisition of expertise in simultaneous interpretation training requires the active development and automatization of extreme language control, a skill which overlaps considerably with other domains of executive control. This form of expertise can be expected to be at least partly driven by an increase in automaticity in performing the task, possibly driven by superior management of its multiple component skills.

After training, we expect to observe reductions in the recruitment of regions crucial to SI, due to increased efficiency of these regions during task performance (as proposed by Ericsson et al., 2006; and see also Kelly and Garavan, 2005). We are particularly interested in the striatum, since it has been shown to be important in polyglot language control. If changes to striatal recruitment are found as a function of training, this will provide evidence for the importance of the striatum in extreme language control.

Methods

Simultaneous interpretation trainees were scanned using fMRI and structural MRI before and after a 15-month intensive training program. Control participants, who were students in other, non-linguistic fields of study, and who had no previous formal experience in interpretation or translation, were also scanned at the same time intervals, using the same paradigms. The experimental procedure was approved by the local research ethics committee (Geneva University Hospital, reference number: 09-161). Participants gave informed consent and were free to withdraw from the experiment at any time, in accordance with the declaration of Helsinki.

Participants

19 trainee simultaneous interpreters (8 female, mean age at time 1: 25 years, range: 22–32 years, mean interval between scans: 1 year, 2 months) and 16 multilingual controls (9 female, mean age at time 1: 24 years, range: 20–33 years, mean interval between scans: 1 year, 2 months) participated in the study. The intervals between scans were not significantly different for the two groups (unpaired t-test, equal variances not assumed[1], $t_{(16.29)} = 0.076$, $p = .945$)[2]. All participants reported a high level of proficiency in a minimum of three languages (ranges: 3–8), and fluently (if not natively) spoke English or French. Participants were asked to report the languages they speak and/or have studied, the age of acquisition of these, and to provide their own assessment of their fluency in each. Following the method described in Golestani et al. (2011) and Hervais-Adelman et al. (2014), a continuous language experience score was then calculated for each individual as follows:

---

[1] Equality of variances for all t-tests was determined using Levene's test of equality of variances.

[2] Here and throughout the paper, t-tests are two-tailed, unless otherwise stated.

knowledge of each language was weighted according to the age at which it had been learnt (higher weight for languages learnt earlier in life) and according to the fluency with which it was spoken (higher weight for more fluently spoken languages). The following weights were used: 1) proficiency: not fluent = 1, somewhat fluent = 2, moderately fluent = 3, quite fluent = 4, very fluent = 5, native = 6; 2) age of acquisition: ages ≥21 years old = 1, ages 13–20 years old = 2, ages 7–12 years old = 3, ages 1–6 years old = 4, at birth = 5. Although there is uncertainty about the validity of self-report as a source of proficiency data, other authors have noted the high correlation between such metrics and objective measures (cf. Marian et al., 2007; see also Morales et al., 2013; Signorelli et al., 2012). There was no significant difference between the groups in terms of their language experience scores (unpaired t-test, equal variances assumed, $t_{(33)} = 1.281$ $p = 0.209$), numbers of languages reported (unpaired t-test, equal variances assumed, $t_{(33)} = 1.055$, $p = 0.299$), age of acquisition of the source language[3] (unpaired t-test, equal variances not assumed, $t_{(23.652)} = 1.118$, $p = 0.275$), or age (unpaired t-test, equal variances assumed, $t_{(33)} = 0.707$, $p = 0.485$). Participants' language and demographic data are presented in Inline Supplementary Table S1. Statistical comparisons of the groups were carried out in SPSS Statistics (IBM Corp. Released 2013. IBM SPSS Statistics for Mac, Version 22.0. Armonk, NY: IBM Corp.).

Inline Supplementary Table S1 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.03.072.

Simultaneous interpretation training

The trainee participants were enrolled in the Master's program in conference interpreting at the Faculty of Translation and Interpretation of the University of Geneva. The course consists of 15-months of intensive training, spread over three 14-week semesters. A first semester in consecutive interpretation (interpretation that commences once the speaker has finished), is followed by two semesters focusing on simultaneous interpretation (interpretation that is more or less synchronized with the original speech). Trainees undergo continuous assessment. Although not all of the participants passed all their final exams at their first attempt, all maintained a standard of performance throughout the course that was at or above the expected level for a trainee at that stage of the training program. The control participants had no formal interpretation or translation experience.

Design and materials

The design employed was identical at both time-points, and is described in (Hervais-Adelman et al., 2014). Participants were scanned using a Siemens 3 T Trio MRI scanner with a 12-channel head-coil. Stimuli were presented through MRI-compatible MR ConFon electrodynamic headphones. A sparse imaging procedure (Hall et al., 1999) was employed, enabling us to both sample the peak of the haemodynamic response function, and to present stimuli and record responses during silent intervals between image acquisitions.

Stimuli consisted of two sets of sentences (A and B), and each set included 156 French and 156 English sentences. These were grammatically simple, with no more than one embedded clause. Sentences were grouped into quartets composing a four-sentence scenario, allowing participants to generate context-based expectation about the content of subsequent sentences. The thematic content of the French and English sets was highly overlapping, though non-identical, due to matching constraints. Sentences were recorded by a bilingual male speaker of Southern British English and French. Participants heard sentences exclusively in one of the two available languages. Sentence quartets were presented in random order, and were randomly allocated to conditions. The order of presentation of sentence sets was counter-

balanced over experimental sessions, such that participants were always presented with a new set of sentences during their second experimental session (i.e. A then B or B then A). Participants were free to choose which language they would hear, and this remained the same between sessions (see Inline Supplementary Table S1 for the distribution of the participants' choices). The source language was always very well mastered by the participants, who were all enrolled on University programs that required them to have a high level of English or French (depending upon the language of instruction of their particular courses), for two early bilingual participants the source language was an L1 (see Inline Supplementary Table S1 for details).

Matching of stimuli between sets and between languages was carried out based on properties of the quartets of sentences (rather than individual sentences). We initially generated and recorded over 90 quartets in French and in English, from which we retained two matched sets of 39 per language. Matching was based on the means of the log10 (frequency per million) of the content words (i.e. excluding articles, auxiliary verbs, prepositions, conjunctions, and pronouns), number of words, total number of syllables and duration of the recording determined for every sentence. Frequencies were retrieved from the entries for subtitles (which better reflect current linguistic usage than frequency estimates drawn from written corpora: New et al., 2007) in the SUBTLEXus database (Brysbaert and New, 2009) for the English sentences and from the Lexique 3.7 database (New et al., 2001) for the French. Matching was carried out using an iterative procedure implemented in Match (van Casteren and Davis, 2007) to produce sets that were initially matched within language. These properties are presented in Inline Supplementary Table S2. These properties were then compared between the lists using a multivariate ANOVA (four measures: frequency, number of words, number of syllables, duration, four levels: French list A and B, English List A and B). A small but significant difference was found in the mean frequency ($F_{(3,152)} = 3.399$, $p = .018$, partial eta-squared = .063). Post-hoc pairwise comparisons revealed that this was driven by a difference between French list A and English list A. However, both of these means are in the top decile of the frequency distribution within the relevant database. Since the order of presentation of the lists was counterbalanced over sessions and there is an approximately even distribution of participants who heard French vs English stimuli, this difference should not be a source of systematic bias in our data. There were no other significant differences between the sets of sentences. For example stimuli, please see Inline Supplementary Table S3.

Inline Supplementary Tables S2 and S3 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.03.072.

At the beginning of each quartet, participants were presented with an on-screen cue for the duration of the preceding image acquisition, consisting of an instruction: "Listen", "Shadow" or "Interpret", which they carried out for the following four sentences. Conditions were pseudo-randomly ordered so that the same instruction was never presented more than three times in succession.

In the "Listen" condition participants were asked simply to listen attentively to the sentences being presented, and to make no response. In the "Shadow" condition, they were asked to repeat the sentences they heard, beginning their response as soon as possible after the onset of each sentence (i.e. to speak while simultaneously hearing each sentence). In the "Interpret" condition they were instructed to simultaneously interpret the content of the sentences they heard into their most fluent language (this was almost invariably their native language, see Inline Supplementary Table S1 for details). Again, participants were instructed to begin their responses as quickly as possible. Verbal responses were recorded for off-line qualitative analysis.

Prior to the experiment, participants underwent brief training inside the scanner, during which they were familiarized with the tasks and the speaker's voice. The experimenters monitored performance during training to ensure that participants were capable of carrying out the interpretation and the shadowing tasks.

---

[3] The language from which the participants interpreted, which was either French or English, see Design and materials section for details.

fMRI data acquisition and processing

In each scanning session, 195 functional volumes were acquired, spread over three runs that lasted approximately 10 min each (T2*-weighted EPI, 36 ∗ 3.2 mm slices, 20% interslice gap, 3.2 ∗ 3.2 mm in-plane resolution angled away from the eyes to prevent ghost-artifacts from aliasing of eye-movements, TA: 2.1 s, TR: 9 s, TE: 80 ms). A total of 13 quartets were presented per condition over the course of each experimental session, resulting in 52 trials per condition. Sentences were presented between scans, and every quartet was succeeded by a single null event in which no stimuli were presented, yielding 65 scans per run. Within each of the three runs, 13 different quartets of sentences were presented. Each run contained four occurrences of each of two conditions and five of the third. The allocation of these was constrained such that each of the conditions could occur five times in only one, randomly-assigned run. Stimuli were presented at a comfortable listening level, adjusted for each participant.

High-resolution (1 mm ∗ 1 mm ∗ 1.2 mm voxels) T1-weighted anatomical images were also acquired for each participant.

Preprocessing and analysis of the data was carried out in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) running in Matlab (Version 2012a, released 2012, The Mathworks, Natick, MA). Data preprocessing for each subject's scans at each time point included: 1) a two-pass rigid-body realignment of each EPI volume to the first in the session, and then to the mean of the realigned images, 2) coregistration of the structural volume to the mean EPI image, 3) normalization of the structural scan to a standard template (SPM8's single-subject T1 template) using the SPM8 unified segmentation routine, 4) normalization of all EPI volumes by application of the deformation parameters estimated in the normalization step, and reslicing the images to 3 mm ∗ 3 mm ∗ 3 mm voxels, and 5) spatial smoothing using a Gaussian kernel of 8 mm full-width half maximum height.

The impact of training was then examined using two different analysis methods, a univariate analysis of variance, and multivariate pattern classification, described below.

Univariate analysis of variance

A first level fixed-effects analysis was carried out using a general linear model (GLM) for each participant in which every scan was coded for condition (Listen, Shadow, or Interpret), and null events were left unmodeled. The duration of sentences was included as a regressor in the design matrix. Each run was modeled separately and the effect of block was coded separately. Each event was modeled using a single finite impulse response function in SPM8. Six parameters were appended in each run to code for the effects of movement (x, y and z translations and x, y and z rotations derived from the rigid realignment step of the pre-processing). A high-pass filter (cutoff 128 s) and AR1 correction for serial autocorrelation were applied.

Contrast images for the pairwise comparisons of Shadow–Listen and Interpret–Listen were produced for each participant. The group data were then analyzed using a 3-way mixed ANOVA with one between-subjects factor ("Group" with two levels: Control or Interpreter) and with two between-subjects factors ("Condition" with two levels: Shadow or Interpret, and "Time" with two levels: "Time 1" and "Time 2"), implemented in GLM_Flex (MacLaren & Schultz; http://nmr.mgh.harvard.edu/harvardagingbrain/People/AaronSchultz/GLM_Flex.html). In order to answer the question addressed by this experiment of whether there was training-related brain functional plasticity in the interpretation trainees and during the interpretation task, we specifically sought the three-way group-by-condition-by-time interaction. Analyses were conducted at a statistical threshold of $p < .005$ (uncorrected for multiple comparisons) with a cluster extent threshold of 5 voxels.

Multivariate pattern classification

In addition to the univariate analysis described above, we conducted a complementary multivariate analysis, in which we examined the overall patterns of differences in activation that were elicited by simultaneous interpretation from the first to the second time-point.

In order to render this problem amenable to investigation by relatively simple pattern classification techniques, we elected to render it a two-class problem by examining the within-subject maps of the between-scan *difference* in BOLD response during simultaneous interpretation vs baseline.

Individual difference maps were computed for each subject by subtracting the T maps for the Interpretation vs Baseline contrast at Time 1 from those at Time 2. These difference maps were labeled as coming from the trained participants or the control participants, and submitted to a multivariate classification procedure using the PRoNTo toolbox (Schrouff et al., 2013), running in Matlab. A linear support vector machine (SVM) was trained to classify the difference maps as originating either from the interpreters or the controls, using a 35-fold leave one subject out procedure. A covariate of the participants' language experience score was also included. The analysis was confined to regions in the brain that had showed significant differences (at a threshold of $p < .001$ uncorrected for multiple comparisons) in activation for any of the conditions (passive listening, shadowing or simultaneous interpretation) compared to baseline at Time 1 or Time 2. A compound mask was created on the basis of these, containing a total of 21,942 voxels (a volume of approximately 592 cm³). This masking was intended to ensure that any patterns that emerged would be attributable to voxels that were implicated in the task. In order to rule out the possibility that results might be due to factors unrelated to the training or the tasks, the same multivariate pattern classification analysis was carried out using a random selection of 21,942 voxels drawn from the rest of the brain. Balanced classification accuracy, weight maps for the SVM and the receiver operating characteristic (ROC) curve were derived and are reported below.

Multivariate pattern classification necessarily produces a result based on all the data in the analyzed region, ascribing a weight for the linear classifier to each and every voxel. The combination of all of these defines the optimal hyperplane of the classifer. Thus, description of the results in a regionwise manner analogous to that typically used in reporting univariate results is somewhat difficult, and potentially even inappropriate. Nevertheless, in order to be able to more easily report, describe and visualize the multivariate patterns we derived, we carried out permutation testing on the weight map of the classification as follows: we repeated the multivariate classification procedure 1000 times, with differently randomly-reassorted labels. A null distribution of weights was derived from the set of permuted maps, and the voxelwise probability of obtaining the weight value generated by the correctly-labeled classifier was estimated. We use this map in order to limit our description of the regions that contribute to the pattern of change to those that are most likely to be reliably implicated in discriminating the groupwise patterns of change between Time 1 and Time 2.

Behavioral responses

Participants' verbal responses during interpretation and during shadowing in the scanner were recorded and checked to ensure compliance with the instructions. Unfortunately, due to equipment failure, recordings from some participants at the second time-point were unintelligible (data-sets were available for 18 interpreters and 12 controls). Responses were scored by a panel of accredited professional simultaneous interpreters trained to work in the language combinations used during the interpretation and shadowing tasks by the participants (one rater per language). Raters were asked to evaluate shadowing responses on a binary scale to indicate failure to comply with instructions (score of 0) and compliance (score of 1). They were asked to assess the interpretation trials on a five-point scale as follows: 0 = no output, 1 = one content word, 2 = two content words (minimally a subject and object), 3 = three content words or more to make a meaningful interpretation, 4 = complete interpretation.

There were a number of response recordings for which the speech content was inaudible, but the presence of a speech envelope could be detected. These were included in an analysis of the response latencies for trials on which a response was produced. Onsets were automatically detected using an in-house procedure implemented in Matlab, based on Runword (Kello and Kawamoto, 1998). A subset for which onsets could not automatically be detected was estimated by hand in Praat (Boersma and Weenink, 2011). Responses were averaged by subject for each condition at each time-point. Partial data were retrieved for 34 of the 35 participants (15 controls and 19 interpreters)[4].

Results

Performance measures

Mean shadowing performance was high in both groups at both time-points, and interpretation performance was very good in the interpreter group and satisfactory in the control group (see Table 1). The effect of group on performance on each condition at each time-point was compared using Mann–Whitney U tests (appropriate for ordinal data). These showed that there was no significant difference between the groups in shadowing performance at either time-point (Time 1: $p = .698$; Time 2: $p = .223$), while there were significant differences for the interpreting condition before ($p = .004$) and after training ($p = .004$), with the trainees outperforming the controls. These differences are not completely unexpected: the interpretation performance of the trainees might be predicted to be better than controls, even before training, by virtue of the fact that trainees are a self-selected group whose pre-training skills met the stringent criteria applied during the written and oral admission exams for the Master's program in conference interpreting. These baseline differences are nonetheless not of direct concern for our study, since we are expressly interested in detecting longitudinal changes in the neural underpinnings of SI in the trained group, in the absence of such changes in the control group.

Training-related changes in interpreting performance were assessed using Friedman's two-way analysis of variance by ranks, and showed that there was a significant improvement in the interpreting condition for the trainee interpreters ($p = .001$) and no change in the controls ($p = .084$). There were no training-related changes in shadowing performance (Interpreters: $p = .155$; Controls: $p = 1$).

Response latencies were also analyzed in order to ensure that potential functional imaging changes over time between groups, in either condition, might not be attributed to different response onsets. We used a 3-way repeated-measures analysis of variance, with within-subjects factors of timepoint (two levels: Time 1, Time 2) and condition (two levels: Shadowing, Interpreting), and a between-subjects factor of group (two levels: Control, Interpreter). There was a significant main effect of session ($F_{(1,32)} = 10.723$, $p = .003$, partial eta-squared = .251), indicating that participants responded more rapidly in the second session (mean onset in session 1 = 1.64 s, SD = 0.34, session 2 = 1.48 s, SD = 0.29). There was also a significant main effect of condition ($F_{(1,32)} = 147.556$, $p < .001$, partial eta-squared = .822), which showed that participants were quicker to respond to shadowing trials than to interpreting trials (mean onset for Shadowing trials = 1.36 s, SD = 0.18, mean onset for Interpreting trials = 1.76 s, SD = 0.31). We wanted to ensure that there was no two-way interaction between group and session, which would reflect a differential change in response latency between groups, in our data. We also wanted to ensure that there was no three-way interaction of group by session by condition, which would indicate a significant change in response latency as a function of condition, by group, between sessions. Neither of these interactions was significant (session by group interaction: $F_{(1,32)} = 0.27$, $p = .609$;

session by group by condition interaction: $F_{(1,32)} = 0.14$, $p = .710$), suggesting that any differences in the functional imaging data cannot be attributed to different response onsets.

Univariate analyses

We specifically sought to reveal brain areas where the response during simultaneous interpretation had changed as a result of training, which is most efficiently represented as a three-way time-by-group-by-condition interaction. Such an interaction was observed in the left inferior frontal gyrus, right caudate nucleus, right middle cingulate cortex, left paracentral lobule, left precentral gyrus and left postcentral gyrus (Fig. 1 and Table 2). In all of these regions, barring the left paracentral lobule, the three-way interaction appears to be driven by decreased activation during interpretation in the trainee group (see interaction plots in Fig. 1).

In order to better characterize the nature of the three-way interaction in the above six clusters, we extracted the individual subjects' data from the peak voxel of each cluster and submitted them to post-hoc testing using SPSS. In order to identify regions in which there was an impact of training, we first sought a group-by-time interaction in the interpretation condition. We conducted a two-way mixed MANOVA with the within-subjects factor of time and the between-subjects factor of group on the data from the six clusters. This analysis identified a significant time-by-group interaction in the right caudate nucleus ($F_{(1,33)} = 4.217$, $p = .048$, partial eta-squared = .113) and a marginally-significant interaction in the left postcentral gyrus ($F_{(1,33)} = 3.952$, $p = .055$, partial eta-squared = .107). Regions that did not show a significant interaction were not further tested, and will not be further discussed. Post-hoc pairwise comparisons revealed that interpretation training results in a significant decrease in the response of the right caudate nucleus during interpretation, in the interpreters ($t_{(18)} = 2.552$, $p = .02$).

In order to determine whether the three-way interaction in the caudate nucleus was specifically driven by changes in the interpretation condition, we also tested for a time-by-group interaction during the shadowing condition in this region. This interaction was not significant ($F_{(1,33)} = 1.633$, $p = .21$).

Multivariate pattern classification analyses

The trained classifier had a balanced classification accuracy of 68.6%. Permutation testing (5000 permutations) established that this classifier was significantly better than chance ($p = .0346$). The receiver operating characteristic (ROC) curve was also calculated and the area under the curve was estimated at 0.66, indicating a moderately sensitive classification function. The control classifier had a balanced classification accuracy of 45.7%, permutation testing (5000 permutations) revealed that this was not significantly better than chance ($p = .6258$). The successful classification of the data indicates that the changes in brain responses, during the simultaneous interpretation condition, of the trained group and the controls differed sufficiently for the SVM to be able to reliably separate the groups. The weight map derived by the SVM is displayed in Fig. 2. The map necessarily comprises all those voxels that were contained in the initial mask.

In order to simplify the description of the pattern and to focus on the voxels with the greatest contribution to the discrimination, we thresholded the weight map to include the top 5% of positively-

Table 1
Mean ratings of behavioral performance (standard deviations in brackets) of 16 interpreters and 12 controls for whom data were available.

	Interpreters		Controls	
	Time 1	Time 2	Time 1	Time 2
Shadowing (max = 1)	0.94(0.05)	0.96(0.06)	0.91(0.14)	0.91(0.10)
Interpreting (max = 4)	3.38(0.21)	3.65(0.15)	2.75(0.62)	3.05(0.63)

---

[4] A total of 3911 onsets were found for the interpreters (99% of the total number of recordings) and a total of 2709 for the controls (87% of the expected responses for the 15 participants for whom recordings were available).
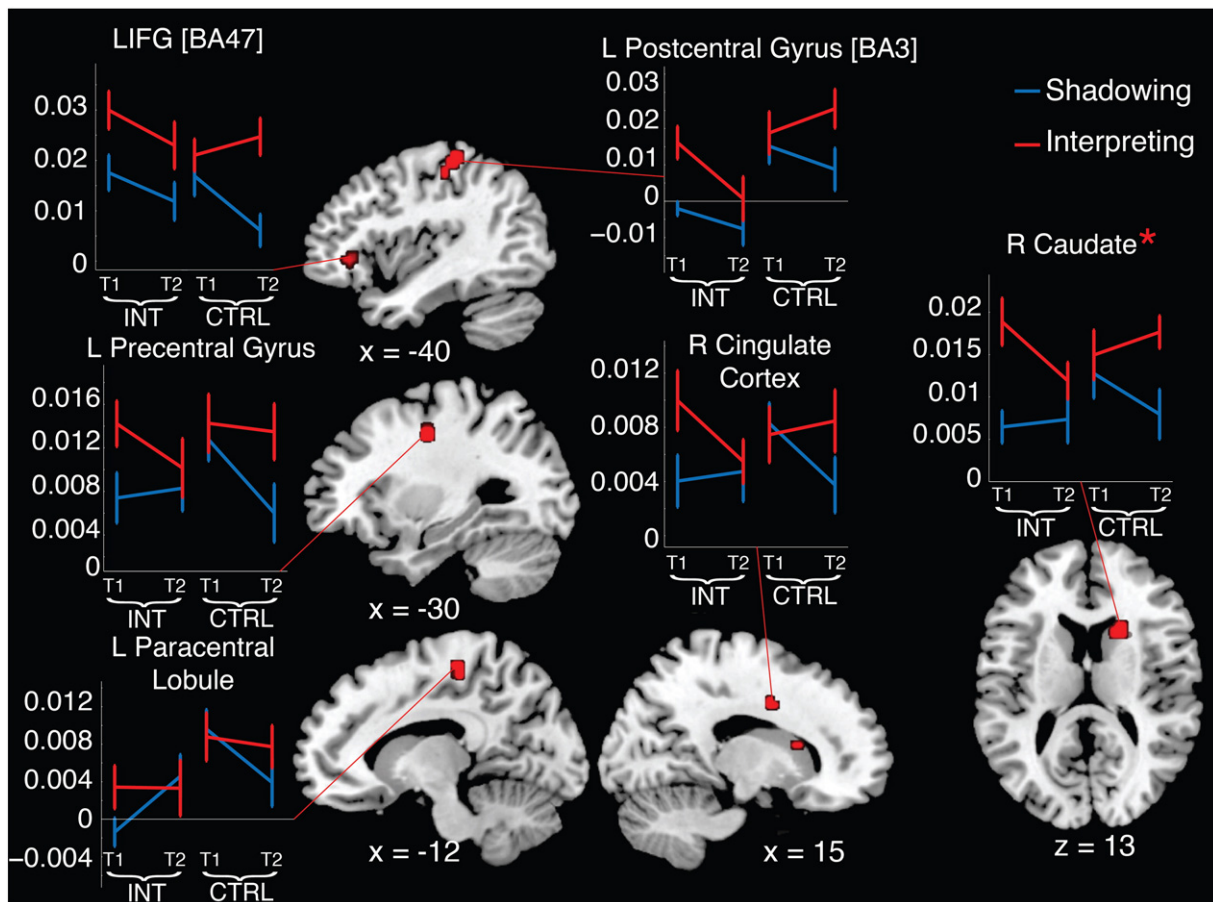
Fig. 1. Three-way interaction of group, time and condition on BOLD response, showing regions within the search volume thresholded at $p < 0.01$, with a cluster extent threshold of 5 voxels, projected on a canonical single-subject brain. Interaction plots represent the mean contrast estimate for each condition vs the silent baseline (in arbitrary units) at the peak voxel of the indicated clusters at each time point (T1 = pre-training, T2 = post-training) for each group (INT = Simultaneous interpretation trainees, CTRL = Controls). Error bars represent standard errors of the mean (corrected to be appropriate for repeated-measures comparisons, as described by Loftus and Masson, 1994). The asterisk indicates a region in which there is a significant effect of training (the three-way interaction is driven by the pre- vs post-training difference in the BOLD response during simultaneous interpretation in the trained participants).

weighted and the top 5% of negatively-weighted voxels. These are highlighted in Fig. 2 and listed in Table 3. The subregions of the pattern in which the contributions to the classifier are greatest are distributed bilaterally and comprise subcortical regions (including the thalamus and the caudate nuclei), the cerebellum, cortical regions associated with speech (the left inferior frontal gyrus, posterior superior temporal regions), hearing (superior temporal cortices) as well as regions involved in movement (sensorimotor and supplementary motor areas). However, since a multivariate pattern is by definition holistic, we will abstain from discussing these findings in detail, save to point out the overlap between regions revealed in the univariate analysis and the multivariate classification. Crucially, our latter analysis demonstrates

that there is a reliable, distributed pattern of change in the brain response during simultaneous interpretation that occurs as a result of training in simultaneous interpretation. We feel it is important to highlight this, as we do not wish to claim that modifications to the response of the caudate nucleus alone are responsible for the considerable superiority in the quality of simultaneous interpreting produced by a trained individual compared to a novice.

Discussion

The results of this study demonstrate that training in simultaneous interpretation induces functional cerebral plasticity in the caudate

Table 2
Clusters showing significant three-way interaction of group by time by condition thresholded at $p < .005$. Bold rows indicate peak voxel in a cluster containing subpeaks. Anatomical labels were retrieved using the AAL [automatic anatomical labeling (Eickhoff et al., 2009)] and BA (Brodmann area) templates provided with the MRIcron software package (http://www.mccauslandcenter.sc.edu/mricro/mricron/ Rorden and Brett, 2000). Abbreviation: Hem = Hemisphere.

| Region name	Hem	MNI co-ordinates (mm)			p (uncorrected for multiple comparisons)	Cluster size (voxels)
		x	y	z		
Inferior frontal gyrus [BA47]	**Left**	**−42**	**32**	**−2**	**<.001**	**9**
Caudate nucleus	**Right**	**21**	**14**	**13**	**<.001**	**16**
Middle cingulate cortex [BA24]	**Right**	**15**	**2**	**37**	**<.001**	**13**
Middle cingulate cortex	Right	18	−7	43		
Paracentral lobule [BA4]	**Left**	**−12**	**−28**	**58**	**<.003**	**7**
Precentral gyrus [BA6]	**Left**	**−30**	**−10**	**43**	**<.003**	**12**
Postcentral gyrus [BA3]	**Left**	**−39**	**−31**	**58**	**<.003**	**13**
Postcentral gyrus	Left	−39	−25	52		
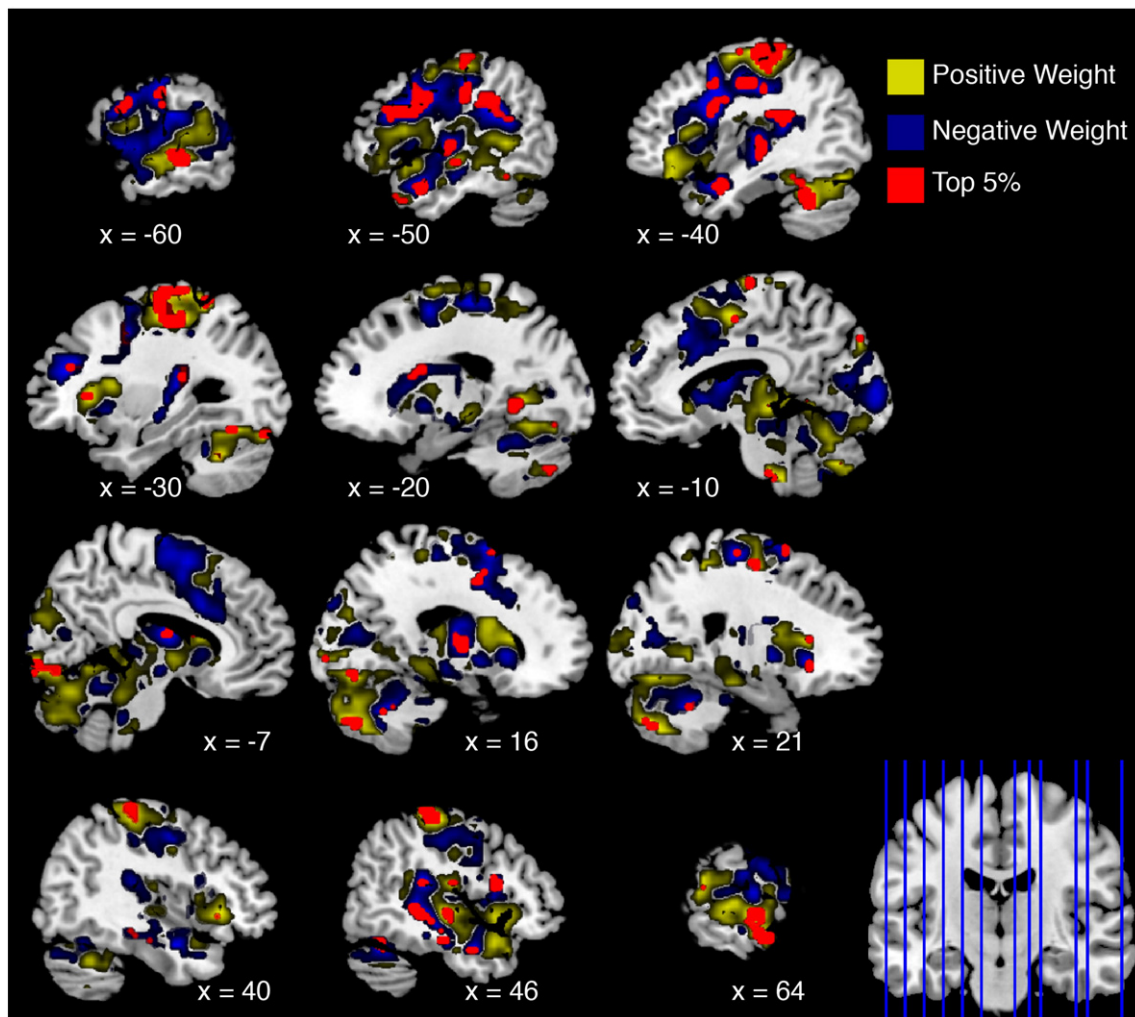
Fig. 2. The weight map representing the optimal hyperplane determined by the linear SVM classifier for discrimination between the changes in functional responses to simultaneous interpretation exhibited by simultaneous interpretation trainees and the control participants, projected on a canonical single-subject brain. Blue indicates positive weights, yellow denotes negative weights, the sign of the weighted sum of an image multiplied by the weights indicates to which group it is ascribed by the classifier. Red highlights those voxels in which the magnitude of the classifier weight is in the top 5% of positive or top 5% of negative values. Planes of section are shown on the coronal slice.

nucleus, which is a key structure in multilingual and executive control, and in a wider distributed network. Univariate analyses indicated that there is decreased engagement of the right caudate nucleus during simultaneous interpretation after training, consistent with existing studies of training-induced functional plasticity (Kelly and Garavan, 2005), and with the notion that expertise and concurrent increases in automaticity of performance lead to reduced use of cerebral resources (Ericsson et al., 2006). The increase in automaticity in task performance in our participants was reflected in interviews after the second scan, in which several of the trainee interpreters indicated that they were no longer aware of the fact that they were speaking a different language to the one they were hearing when carrying out simultaneous interpretation, and that they rather had a sensation of simply repeating what they heard. Multivariate pattern classification of maps of pre- vs post-training activation differences during simultaneous interpretation demonstrated that there was a significant change in task-specific responses, sufficiently reliable as to be able to discriminate the trained group from the control participants.

Although the size of the effects we observe is relatively small, it should be noted that simultaneous interpretation is a highly complex task, loading heavily on a wide variety of cognitive and linguistic skills, allowing for a range of strategies to ensure successful results. The regions in which we find reliable changes in activation as a function of

training are therefore likely to be the most stable nodes (in terms of neuroanatomical loci) implicated in the task.

Existing neuroimaging work on bilingual language control has strongly implicated the dorsal striatum (reviewed by Friederici, 2006; and by Hervais-Adelman et al., 2011), with investigations of language switching having revealed a particular role for the caudate nucleus, bilaterally (Abutalebi and Green, 2008; Garbin et al., 2010; Luk et al., 2012; Moritz-Gasser and Duffau, 2009). Neuroimaging studies of translation have also implicated basal ganglia structures including the putamen and caudate nucleus (Price et al., 1999), and the external globus pallidus (Lehtonen et al., 2005). Neuropsychological reports of cases of subcortical polyglot aphasia converge with findings of neuroimaging studies, with evidence for a role for the dorsal striatum in language control. In these, damage to the putamen and caudate has been seen to provoke uncontrolled language switching and mixing (Abutalebi et al., 2000, 2013; Aglioti et al., 1996; Moretti et al., 2001), and a recent case study has shown that direct stimulation of the caudate nucleus has a detrimental impact on language selection (Wang et al., 2012).

Hervais-Adelman et al. (2014) previously proposed that during simultaneous interpretation, the caudate nucleus serves to control access to two lexico-semantic systems in response to the ongoing demands of interpretation, which requires rapid and concurrent access to two languages. Reduced activation in this region, as revealed in the univariate

Table 3
The top 5% of the positive and top 5% of the negative weights of the SVM classifier that distinguishes between patterns of Time 1 vs Time 2 change in response during simultaneous interpretation. This table serves exclusively to highlight those regions that contribute most to the map — the full set of weights contributes to the classification of the data. Anatomical labels retrieved as for Table 2. Abbreviation: Hem = Hemisphere.

Region name [Brodmann area]	Hem	MNI co-ordinates (mm)			Voxels in cluster	Max absolute weight in cluster	Mean weight in cluster
		x	y	z			
Middle temporal gyrus [BA21]	L	−66	−28	−2	37	−0.036	−0.019
Postcentral gyrus [BA3]	L	−42	−31	61	148	−0.029	−0.019
Superior temporal plane	L	−39	−37	22	72	0.029	0.019
Inferior temporal gyrus [BA37]	R	51	−61	−23	10	0.025	0.019
Inferior frontal gyrus pars opercularis [BA44]	R	48	11	16	8	0.025	0.02
Superior temporal sulcus [BA21]	R	66	−10	−2	36	−0.024	−0.018
Paracentral lobule [BA6]	L	−12	−19	76	12	−0.024	−0.019
Superior temporal sulcus	R	48	−19	−5	12	−0.023	−0.019
Superior temporal plane [BA21]	R	45	−37	−2	11	0.023	0.019
Precentral sulcus [BA6]	L	−33	5	43	140	0.022	0.017
Anterior superior temporal sulcus [BA36]	L	−39	2	−23	25	0.022	0.017
Inferior temporal sulcus [BA20]	R	51	−25	−17	21	0.022	0.019
Anterior lateral thalamus	R	15	−7	7	9	0.022	0.019
Postcentral sulcus	L	−60	−19	34	8	0.022	0.017
Fusiform gyrus [BA37]	L	−42	−46	−20	36	−0.021	−0.017
Anterior lingual gyrus [BA19]	L	−18	−55	−5	16	−0.021	−0.017
Posterior lingual gyrus [BA19]	L	−33	−85	−17	8	−0.021	−0.018
Superior temporal sulcus [BA22]	L	−54	−16	−8	7	−0.021	−0.017
Postcentral gyrus [BA3]	R	45	−28	58	26	−0.02	−0.017
Lingual gyrus [BA18]	R	12	−79	−11	22	−0.02	−0.017
Medial anterior Heschl's gyrus	L	−39	−25	1	17	0.02	0.017
Central sulcus	L	−36	−16	40	16	0.02	0.017
Cerebellum crus II	R	15	−79	−41	15	−0.02	−0.017
Posterior superior temporal plane [BA42]	R	54	−28	13	10	0.02	0.017
Postcentral sulcus [BA3]	L	−45	−25	34	21	0.019	0.017
Cerebellum lobule VI	R	30	−46	−35	21	0.019	0.017
Middle Heschl's gyrus	L	−48	−13	1	9	0.019	0.017
Fusiform gyrus [BA19]	L	−27	−67	−14	7	−0.019	−0.016
Caudate nucleus	R	9	8	13	6	−0.019	−0.017
Middle temporal pole [BA21]	R	54	11	−29	14	−0.018	−0.016
Superior parietal lobule [BA2]	L	−33	−46	64	8	−0.018	−0.017
Cerebellum lobule VII	L	−18	−73	−41	8	−0.018	−0.016
Caudate nucleus	L	−18	5	22	7	0.018	0.017
Brainstem [ventral pons/superior olivary complex]	L	−9	−28	−44	6	−0.018	−0.016
Supplementary motor area [BA6]	R	12	2	49	5	0.018	0.017
Anterior insula/inferior frontal gyrus pars orbitalis [BA47]	L	−27	26	4	7	−0.016	−0.016
Inferior frontal sulcus [BA46]	L	−33	41	22	6	0.016	0.016
Superior frontal sulcus/motor cortex [BA6]	R	18	11	67	5	0.016	0.016

analysis, suggests that training reduces the load placed on this network to effectively carry out the task. Previous imaging studies of simultaneous interpretation in highly-experienced interpreters failed to find evidence for the involvement of the caudate nucleus during simultaneous interpretation (Rinne et al., 2000; Tommola et al., 2000). Although these absences provide only circumstantial evidence, it is plausible that its engagement may further decrease with increasing expertise. Future studies investigating interpreters with different degrees of expertise could shed light on this matter.

An intriguing aspect of our results is the lateralization of the training effect to the right caudate nucleus. The evidence from both lesion studies (for example Abutalebi et al., 2000, 2009) and functional imaging of bilingual language control (Abutalebi and Green, 2008; Garbin et al., 2010; Klein et al., 1994, 1995, 2006; Lehtonen et al., 2005; Price et al., 1999) has tended to suggest that this function of the basal ganglia is left-lateralized. However, a meta-analysis of neuroimaging studies of language switching indicates that both right and left striatal structures are implicated in language switching (Luk et al., 2012). Furthermore, Hervais-Adelman et al. (2014) reported bilateral caudate involvement in simultaneous interpreting, with an apparent rightward lateralization, and a number of experiments have also implicated the right striatum in translation tasks (e.g. Abutalebi and Green, 2007; Price et al., 1999). In addition, previous brain structural studies in experienced simultaneous interpreters have shown differences between interpreters and controls in white matter properties of the right caudate nucleus (Elmer et al., 2011), and in a related study, a negative correlation was found between

the amount of SI practice and gray matter volume in the caudate nuclei bilaterally in interpreters (Elmer et al., 2011, 2014).

One possible explanation for the apparent lateralization of the effect in the present study is that the left striatum is more directly involved in motor aspects of speech control (as suggested by the evidence from direct electrical stimulation by Gil Robles et al., 2005), which may leave less scope for training-related reductions in its engagement compared to the right striatum. Further careful investigation of the activity in the striatum during monolingual and bilingual speech production tasks is needed to test this notion. Alternatively, although rather more prosaically, it is possible that in the present data, we were more easily able to reveal training-related activation changes in the right than the left caudate nucleus because of the initially higher level of activation in the right at baseline.

The caudate nucleus is not a language-specific structure. It is implicated in a vast range of functions, from motor control to cognitive control (Grahn et al., 2008, 2009; Graybiel, 1995a,b, 1997, 2005; Houk et al., 2007; Houk and Wise, 1995; Kropotov and Etlinger, 1999; Middleton and Strick, 2000a,b). The precise nature of its role in language is not entirely clear, but a compelling proposal by Kotz et al. (2009) suggests that the role of the basal ganglia in speech relates to sequencing and prediction, a concept also eloquently set out by Saint-Cyr (2003) with respect to the role of the basal ganglia in information processing. These aspects of language processing are of fundamental importance to simultaneous interpreting, particularly between asymmetric language pairs, when the interpreter cannot afford to wait until an entire idea unfolds in the

original discourse, but must predict the outcome so as to relieve the load that would otherwise be placed on rehearsal in working memory.

The fact that the structures of the dorsal striatum are implicated in multilingual language control is of particular relevance to the now oft-reported "bilingual advantage" on many executive tasks requiring conflict resolution, attentional control, action selection and cognitive flexibility. There is a suggestion that the ongoing exercise of the mechanisms of language control in the bilingual brain serves to enhance domain general cognitive capacities. A possible neuroanatomical locus for this effect is in the basal ganglia, a view propounded by Stocco et al. (2012), who suggest that modifications to the striatum, resulting from the exercise of multilingualism, could be responsible for the cognitive advantages reported in bilingual individuals over their monolingual peers (for reviews see Bialystok, 2009; Bialystok et al., 2012; Diamond, 2010). Our study provides a clear demonstration that deliberate language-management practice induces functional changes in the right caudate nucleus, over and above multilingualism per se.

In the next section we speculate upon the relationship between the caudate nucleus and expertise, not only in interpreting, but also in other domains.

The role of the caudate nucleus in expertise

Functional changes in the caudate nuclei as a result of expertise have been found, cross-sectionally, in a variety of domains. Expert musicians carrying out a bimanual coordination task display less right striatal activation than do non-musicians (Haslinger et al., 2004). Duan et al. (2012) showed that chess masters had increased resting-state functional connectivity between the caudate nuclei and the default mode network. While music and strategic planning are often thought of as principally cognitive tasks, functional plasticity also occurs in the basal ganglia of experts in an apparently unrelated domain: golf. Milton et al. (2007) showed that expert golfers displayed no significant recruitment of the basal ganglia during pre-shot motor planning, whereas novices did. They conclude that this reflects a change to an invariant component of the motor planning network that differs between the two levels of expertise. Expertise-related structural changes have also been reported: expert musicians exhibit decreased gray-matter density in the right caudate as a function of musical expertise (James et al., 2013), and expert interpreters similarly show decreased gray-matter density in the caudate nuclei bilaterally as a function of cumulative expertise (Elmer et al., 2014). The intersection of these disparate domains of expertise in the right striatum suggests that decreased caudate engagement may be a cerebral hallmark of expert performance. While we postulate that during simultaneous interpretation the caudate nucleus serves to coordinate the cognitive subcomponents of the task, it is important to recognize that the training-induced reduction in caudate activity may reflect a general characteristic of expert performance in that expertise depends upon the appropriate management of cognitive resources, in order to respond rapidly and to a considerable extent automatically to the highly variable demands of task performance. With practice, these demands are reduced. The caudate nucleus exhibits strong structural connectivity with frontal brain areas that are implicated in executive functions (Draganski et al., 2008), and is known to play a central role in learning (Beauchamp et al., 2003), memory and motivation (Packard and Knowlton, 2002), as well as representing action–outcome contingencies (Grahn et al., 2008). This multiplicity of roles makes it a plausible hub for coordinating the cerebral subsystems involved in task execution across domains of perceptual, cognitive, and motor performance.

Limitations

As described in the behavioral results, we found group-wise differences in performance before training, differences that may arise from the fact that trainees are a self-selected group whose domain-specific

(i.e. interpretation and related) skills allowed them to enter the SI training program. These baseline differences do not, however, affect the interpretation of our longitudinal, training-related results.

It should also be noted that the task we used may have been rather easy for the trained participants, whose practice involves the simultaneous interpretation of passages of speech of a much more complex nature than the stimuli we produced. Also, we specifically designed our stimuli to minimize the relevance of world knowledge in order to make the task feasible before training and for the control participants, whereas the real-life practice of SI requires the active use of world knowledge, such as acquired through preparatory reading and analysis of subject-matter relevant documentation. It should be noted, however, that one investigation of the utility of prior knowledge for interpreting in professional interpreters showed that neither access to a précis nor to the full-text of a speech improved performance (Anderson, 1994). Nonetheless, it cannot be ruled out that handling world knowledge efficiently is a skill that interpreters develop that contributes to their ability to interpret well, and one to which we were insensitive in this study.

Conclusions

The activation decrease we observe in the right caudate nucleus during simultaneous interpretation after training is indicative of refinement of the network engaged in the task. The changes in the caudate nucleus may be due to increased efficiency in juggling the lexico-semantic sets of two languages required for the execution of simultaneous interpretation. Our findings add weight to the notion that the right striatum occupies a key position in evolving expert performance, as has previously been shown in professional musicians (James et al., 2013), chess players (Haslinger et al., 2004), athletes (Duan et al., 2012) and here, in simultaneous interpreters. This notion is compatible with the wide-ranging roles of the basal ganglia, which include cognitive pattern generation (Milton et al., 2007; Yarrow et al., 2009), action pattern selection and refinement (Graybiel, 1997), and prediction (Houk et al., 2007), and in learning, memory and motivation more generally. Expertise requires the ability to rapidly and dynamically respond to various inputs within the trained domain, and requires the flexible deployment of acquired action repertoires — a role tailor-made for a structure that is profoundly implicated in the control of goal-directed action (Kotz et al., 2009). Our findings provide clear evidence that the deliberate exercise of multilingual language control shapes the brain networks involved in executive control.

Acknowledgments

References

Abutalebi, J., Green, D.W., 2007. Bilingual language production: the neurocognition of language representation and control. J. Neurolinguistics 20 (3), 242–275.

Abutalebi, J., Green, D.W., 2008. Control mechanisms in bilingual language production: neural evidence from language switching studies. Lang. Cogn. Process. 23 (4), 557–582. http://dx.doi.org/10.1080/01690960801920602.

Abutalebi, J., Miozzo, A., Cappa, S.F., 2000. Do subcortical structures control 'language selection' in polyglots? Evidence from pathological language mixing. Neurocase 6 (1), 51–56.

Abutalebi, J., Rosa, P.A., Tettamanti, M., Green, D.W., Cappa, S.F., 2009. Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. Brain Lang. 109 (2-3), 141–156. http://dx.doi.org/10.1016/j.bandl.2009.03.003.

Abutalebi, J., Della Rosa, P.A., Green, D.W., Hernandez, M., Scifo, P., Keim, R., Cappa, S.F., Costa, A., 2012. Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cereb Cortex 22 (9), 2076–2086.

Abutalebi, J., Della Rosa, P.A., Gonzaga, A.K., Keim, R., Costa, A., Perani, D., 2013. The role of the left putamen in multilingual language production. Brain Lang. 125 (3), 307–315. http://dx.doi.org/10.1016/j.bandl.2012.03.009.

Agliotti, S., Beltramello, A., Girardi, F., Fabbro, F., 1996. Neurolinguistic and follow-up study of an unusual pattern of recovery from bilingual subcortical aphasia. Brain 119 (Pt 5), 1551–1564 (5).

Anderson, L., 1994. Simultaneous interpretation: contextual and translation aspects. In: Moser-Mercer, B., Lambert, S. (Eds.), Bridging the Gap: Empirical Research in Simultaneous Interpretation. John Benjamins, Amsterdam.

Aron, A.R., Behrens, T.E., Smith, S., Frank, M.J., Poldrack, R.A., 2007a. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. J. Neurosci. 27 (14), 3743–3752. http://dx.doi.org/10.1523/JNEUROSCI.0519-07.2007.

Aron, A.R., Durston, S., Eagle, D.M., Logan, G.D., Stinear, C.M., Stuphorn, V., 2007b. Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. J. Neurosci. 27 (44), 11860–11864. http://dx.doi.org/10.1523/JNEUROSCI.3644-07.2007.

Beauchamp, M.H., Dagher, A., Aston, J.A., Doyon, J., 2003. Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. NeuroImage 20 (3), 1649–1660.

Bialystok, E., 2009. Bilingualism: the good, the bad, and the indifferent. Biling. Lang. Cogn. 12 (1), 3–11. http://dx.doi.org/10.1017/S1366728908003477.

Bialystok, E., Craik, F.I., Luk, G., 2012. Bilingualism: consequences for mind and brain. Trends Cogn. Sci. 16 (4), 240–250. http://dx.doi.org/10.1016/j.tics.2012.03.001.

Boersma, P., Weenink, D., 2011. Praat: Doing Phonetics by Computer (Version 5.2.23). Retrieved from. http://www.praat.org.

Brysbaert, M., New, B., 2009. Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behav. Res. Methods 41 (4), 977–990. http://dx.doi.org/10.3758/Brm.41.4.977.

Christoffels, I.K., de Groot, A., Kroll, J.F., 2006. Memory and language skills in simultaneous interpreters: the role of expertise and language proficiency. J. Mem. Lang. 54 (3), 324–345.

Colzato, L.S., Bajo, M.T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., Hommel, B., 2008. How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. J. Exp. Psychol. Learn. Mem. Cogn. 34 (2), 302–312. http://dx.doi.org/10.1037/0278-7393.34.2.302.

Costa, A., Sebastian-Galles, N., 2014. How does the bilingual experience sculpt the brain? Nat. Rev. Neurosci. 15 (5), 336–345. http://dx.doi.org/10.1038/nrn3709.

Costa, A., Hernandez, M., Costa-Faidella, J., Sebastian-Galles, N., 2009. On the bilingual advantage in conflict processing: Now you see it, now you don't. Cognition 113 (2), 135–149. http://dx.doi.org/10.1016/j.cognition.2009.08.001.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J.T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D.W., Price, C.J., 2006. Language control in the bilingual brain. Science 312 (5779), 1537–1540.

Daró, V., Fabbro, F., 1994. Verbal memory during simultaneous interpretation: effects of phonological interference. Appl. Linguist. 15 (4), 365–381.

Diamond, J., 2010. The benefits of multilingualism. Science 330 (6002), 332–333. http://dx.doi.org/10.1126/science.1195067.

Draganski, B., Kherif, F., Kloppel, S., Cook, P.A., Alexander, D.C., Parker, G.J., Deichmann, R., Ashburner, J., Frackowiak, R.S., 2008. Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. J Neurosci 28 (28), 7143–7152.

Duan, X., Liao, W., Liang, D., Qiu, L., Gao, Q., Liu, C., Gong, Q., Chen, H., 2012. Large-scale brain networks in board game experts: insights from a domain-related task and task-free resting state. PLoS One 7 (3) e32532.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp. 30 (9), 2907–2926. http://dx.doi.org/10.1002/hbm.20718.

Elmer, S., Meyer, M., Jancke, L., 2010. Simultaneous interpreters as a model for neuronal adaptation in the domain of language processing. Brain Res. 1317, 147–156. http://dx.doi.org/10.1016/j.brainres.2009.12.052.

Elmer, S., Hanggi, J., Meyer, M., Jancke, L., 2011. Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. Hum. Brain Mapp. 32 (12), 2064–2074. http://dx.doi.org/10.1002/hbm21169.

Elmer, S., Hanggi, J., Jancke, L., 2014. Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: insights from simultaneous interpreters. Cortex 54, 179–189. http://dx.doi.org/10.1016/j.cortex.2014.02.014.

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R., 2006. The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, New York, NY, US.

Fabbro, F., Daró, V., 1995. Delayed auditory feedback in polyglot simultaneous interpreters. Brain Lang. 48 (3), 309–319. http://dx.doi.org/10.1006/brln.1995.1013.

Friederici, A.D., 2006. What's in control of language? Nat. Neurosci. 9 (8), 991–992. http://dx.doi.org/10.1038/nn0806-991.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J.C., Rodriguez-Pujadas, A., Belloch, V., Hernandez, M., Costa, A., Avila, C., 2010. Bridging language and attention: brain basis of the impact of bilingualism on cognitive control. Neuroimage 53 (4), 1272–1278.

Gil Robles, S., Gatignol, P., Capelle, L., Mitchell, M.C., Duffau, H., 2005. The role of dominant striatum in language: a study using intraoperative electrical stimulations. J. Neurol. Neurosurg. Psychiatry 76 (7), 940–946. http://dx.doi.org/10.1136/jnnp.2004.045948.

Golestani, N., Price, C.J., Scott, S.K., 2011. Born with an ear for dialects? Structural plasticity in the expert phonetician brain. J. Neurosci. 31 (11), 4213–4220. http://dx.doi.org/10.1523/JNEUROSCI.3891-10.2011.

Grahn, J.A., Parkinson, J.A., Owen, A.M., 2008. The cognitive functions of the caudate nucleus. Prog. Neurobiol. 86 (3), 141–155. http://dx.doi.org/10.1016/j.pneurobio.2008.09.004.

Grahn, J.A., Parkinson, J.A., Owen, A.M., 2009. The role of the basal ganglia in learning and memory: neuropsychological studies. Behav. Brain Res. 199 (1), 53–60. http://dx.doi.org/10.1016/j.bbr.2008.11.020.

Graybiel, A.M., 1995a. The basal ganglia. Trends Neurosci. 18 (2), 60–62 (doi: 016622369593871 T, pii).

Graybiel, A.M., 1995b. Building action repertoires: memory and learning functions of the basal ganglia. Curr. Opin. Neurobiol. 5 (6), 733–741 (doi: 0959-4388(95)80100-6, pii).

Graybiel, A.M., 1997. The basal ganglia and cognitive pattern generators. Schizophr. Bull. 23 (3), 459–469.

Graybiel, A.M., 2005. The basal ganglia: learning new tricks and loving it. Curr. Opin. Neurobiol. 15 (6), 638–644. http://dx.doi.org/10.1016/j.conb.2005.10.006.

Green, D., 1998. Mental control of the bilingual lexico-semantic system. Biling. Lang. Cogn. 1 (2), 67–81.

Green, D.W., Abutalebi, J., 2013. Language control in bilinguals: the adaptive control hypothesis. J. Cogn. Psychol. (Hove) 25 (5), 515–530. http://dx.doi.org/10.1080/20445911.2013.796377.

Grogan, A., Parker-Jones, O., Ali, N., Crinion, J., Orabona, S., Mechias, M.L., Ramsden, S., Green, D.W., Price, C.J., 2012. Structural correlates for lexical efficiency and number of languages in non-native speakers of English. Neuropsychologia 50 (7), 1347–1352.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7 (3), 213–223.

Haslinger, B., Erhard, P., Altenmuller, E., Hennenlotter, A., Schwaiger, M., Grafin von Einsiedel, H., Rummeny, E., Conrad, B., Ceballos-Baumann, A.O., 2004. Reduced recruitment of motor association areas during bimanual coordination in concert pianists. Hum Brain Mapp 22 (3), 206–215.

Hervais-Adelman, A., Moser-Mercer, B., Golestani, N., 2011. Executive control of language in the bilingual brain: integrating the evidence from neuroimaging to neuropsychology. Front. Psychol. 2, 234. http://dx.doi.org/10.3389/fpsyg.2011.00234.

Hervais-Adelman, A., Moser-Mercer, B., Michel, C.M., Golestani, N., 2014. fMRI of simultaneous interpretation reveals the neural basis of extreme language control. Cereb. Cortex http://dx.doi.org/10.1093/cercor/bhu158.

Houk, J.C., Wise, S.P., 1995. Feature article: distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. Cereb. Cortex 5 (2), 95–110.

Houk, J.C., Bastianen, C., Fansler, D., Fishbach, A., Fraser, D., Reber, P.J., Roy, S.A., Simo, L.S., 2007. Action selection and refinement in subcortical loops through basal ganglia and cerebellum. Philos Trans R Soc Lond B Biol Sci 362 (1485), 1573–1583.

James, C.E., Oechslin, M.S., Van De Ville, D., Hauert, C.A., Descloux, C., Lazeyras, F., 2013. Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. Brain Struct. Funct. 219 (1), 353–366. http://dx.doi.org/10.1007/s00429-013-0504-z.

Kello, C.T., Kawamoto, A.H., 1998. Runword: an IBM-PC software package for the collection and acoustic analysis of speeded naming responses. Behav. Res. Methods Instrum. Comput. 30, 371–383.

Kelly, A.M., Garavan, H., 2005. Human functional neuroimaging of brain changes associated with practice. Cereb. Cortex 15 (8), 1089–1102. http://dx.doi.org/10.1093/cercor/bhi005.

Klein, D., Zatorre, R.J., Milner, B., Meyer, E., Evans, A.C., 1994. Left putaminal activation when speaking a second language: evidence from PET. Neuroreport 5 (17), 2295–2297.

Klein, D., Milner, B., Zatorre, R.J., Meyer, E., Evans, A.C., 1995. The neural substrates underlying word generation — a bilingual functional-imaging study. Proc. Natl. Acad. Sci. U. S. A. 92 (7), 2899–2903.

Klein, D., Watkins, K.E., Zatorre, R.J., Milner, B., 2006. Word and nonword repetition in bilingual subjects: a PET study. Hum. Brain Mapp. 27 (2), 153–161. http://dx.doi.org/10.1002/hbm.20174.

Klein, D., Mok, K., Chen, J.K., Watkins, K.E., 2014. Age of language learning shapes brain structure: a cortical thickness study of bilingual and monolingual individuals. Brain Lang. 131, 20–24. http://dx.doi.org/10.1016/j.bandl.2013.05.014.

Kopke, B., Nespoulous, J.-L., 2006. Working memory performance in expert and novice interpreters. Interpreting 8 (1), 1–23.

Kotz, S.A., Schwartze, M., Schmidt-Kassow, M., 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. Cortex 45 (8), 982–990. http://dx.doi.org/10.1016/j.cortex.2009.02.010.

Kropotov, J.D., Etlinger, S.C., 1999. Selection of actions in the basal ganglia-thalamocortical circuits: review and model. Int. J. Psychophysiol. 31 (3), 197–217 (doi: S0167-8760(98)00051-8, pii).

Lehtonen, M.H., Laine, M., Niemi, J., Thomsen, T., Vorobyev, V.A., Hugdahl, K., 2005. Brain correlates of sentence translation in Finnish–Norwegian bilinguals. Neuroreport 16 (6), 607–610 (doi: 00001756-200504250-00018, pii).

Loftus, G.R., Masson, M.E.J., 1994. Using confidence intervals in within-subjects designs. Psychon. Bull. Rev. 1 (4), 476–490.

Luk, G., Green, D.W., Abutalebi, J., Grady, C., 2012. Cognitive control for language switching in bilinguals: a quantitative meta-analysis of functional neuroimaging studies. Lang. Cogn. Process. 27 (10), 1479–1488. http://dx.doi.org/10.1080/01690965.2011.613209.

Marian, V., Blumenfeld, H.K., Kaushanskaya, M., 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. J. Speech Lang. Hear. Res. 50 (4), 940–967. http://dx.doi.org/10.1044/1092-4388(2007/067).

Mechelli, A., Crinion, J.T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R.S., Price, C.J., 2004. Neurolinguistics: structural plasticity in the bilingual brain. Nature 431 (7010), 757. http://dx.doi.org/10.1038/431757a.

Middleton, F.A., Strick, P.L., 2000a. Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res. Brain Res. Rev. 31 (2-3), 236–250.

Middleton, F.A., Strick, P.L., 2000b. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. Brain Cogn. 42 (2), 183–200. http://dx.doi.org/10.1006/brcg.1999.1099.

Milton, J., Solodkin, A., Hlustik, P., Small, S.L., 2007. The mind of expert motor performance is cool and focused. NeuroImage 35 (2), 804–813. http://dx.doi.org/10.1016/j.neuroimage.2007.01.003.

Morales, J., Gomez-Ariza, C.J., Bajo, M.T., 2013. Dual mechanisms of cognitive control in bilinguals and monolinguals. J. Cogn. Psychol. (Hove) 25 (5), 531–546. http://dx.doi.org/10.1080/20445911.2013.807812.

Moretti, R., Bava, A., Torre, P., Antonello, R.M., Zorzon, M., Zivadinov, R., Cazzato, G., 2001. Bilingual aphasia and subcortical–cortical lesions. Percept. Mot. Skills 92 (3 Pt 1), 803–814.

Moritz-Gasser, S., Duffau, H., 2009. Cognitive processes and neural basis of language switching: proposal of a new model. Neuroreport 20 (18), 1577–1580. http://dx.doi.org/10.1097/WNR.0b013e328333907e.

Moser, B., 1978. Simultaneous interpretation: a hypothetical model and its practical application. In: Gerver, D., Sinaiko, H.W. (Eds.), Language, Interpretation and Communication. Plenum Press, New York, pp. 353–368.

Moser-Mercer, B., Lambert, S., Daró, V., Williams, S., 1997. Skill components in simultaneous interpreting. Paper Presented at the Conference Interpreting: Current Trends in Research, Turku.

New, B., Pallier, C., Ferrand, L., Matos, R., 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE. Annee Psychol. 101, 447–462.

New, B., Brysbaert, M., Veronis, J., Pallier, C., 2007. The use of film subtitles to estimate word frequencies. Appl. Psycholinguist. 28 (4), 661–677. http://dx.doi.org/10.1017/S014271640707035x.

Packard, M.G., Knowlton, B.J., 2002. Learning and memory functions of the Basal Ganglia. Annu. Rev. Neurosci. 25, 563–593. http://dx.doi.org/10.1146/annurev.neuro.25.112701.142937.

Price, C.J., Green, D.W., von Studnitz, R., 1999. A functional imaging study of translation and language switching. Brain 122 (Pt 12), 2221–2235.

Rinne, J.O., Tommola, J., Laine, M., Krause, B.J., Schmidt, D., Kaasinen, V., Teras, M., Sipila, H., Sunnari, M., 2000. The translating brain: cerebral activation patterns during simultaneous interpreting. Neurosci Lett 294 (2), 85–88.

Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. Behav. Neurol. 12 (4), 191–200.

Saint-Cyr, J.A., 2003. Frontal-striatal circuit functions: context, sequence, and consequence. J. Int. Neuropsychol. Soc. 9 (1), 103–127.

Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mourao-Miranda, J., 2013. PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics 11 (3), 319–337.

Signorelli, T.M., Haarmann, H.J., Obler, L.K., 2012. Working memory in simultaneous interpreters: effects of task and age. Int. J. Biling. 16 (2), 198–212.

Stavrakaki, S., Megari, K., Kosmidis, M.H., Apostolidou, M., Takou, E., 2012. Working memory and verbal fluency in simultaneous interpreters. J. Clin. Exp. Neuropsychol. 34 (6), 624–633. http://dx.doi.org/10.1080/13803395.2012.667068.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., Brandeis, D., Dierks, T., 2012. Structural plasticity in the language system related to increased second language proficiency. Cortex 48 (4), 458–465.

Stocco, A., Yamasaki, B., Natalenko, R., Prat, C.S., 2012. Bilingual brain training: a neurobiological framework of how bilingual experience improves executive function. Int. J. Biling. 18 (1), 67–92. http://dx.doi.org/10.1177/1367006912456617.

Tommola, J., Laine, M., Sunnari, M., Rinne, J.O., 2000. Images of shadowing and interpreting. Interpreting 5 (2), 147–169.

van Casteren, M., Davis, M.H., 2007. Match: a program to assist in matching the conditions of factorial experiments. Behav. Res. Methods 39 (4), 973.

van Heuven, W.J.B., Schriefers, H., Dijkstra, T., Hagoort, P., 2008. Language conflict in the bilingual brain. Cereb Cortex 18 (11), 2706–2716.

Wang, X., Wang, Y.Y., Jiang, T., Wang, Y.Z., Wu, C.X., 2012. Direct evidence of the left caudate's role in bilingual control: an intra-operative electrical stimulation study. Neurocase 19 (5), 462–469. http://dx.doi.org/10.1080/13554794.2012.701635.

Yarrow, K., Brown, P., Krakauer, J.W., 2009. Inside the brain of an elite athlete: the neural processes that support high achievement in sports. Nat. Rev. Neurosci. 10 (8), 585–596. http://dx.doi.org/10.1038/nrn2672.

Yudes, C., Macizo, P., Bajo, T., 2011. The influence of expertise in simultaneous interpreting on non-verbal executive processes. Front. Psychol. 2, 309. http://dx.doi.org/10.3389/fpsyg.2011.00309.

Yudes, C., Macizo, P., Bajo, T., 2012. Coordinating comprehension and production in simultaneous interpreters: evidence from the articulatory suppression effect. Biling. Lang. Cogn. 15 (2), 329–339. http://dx.doi.org/10.1017/S1366728911000150.

Zou, L., Ding, G., Abutalebi, J., Shu, H., Peng, D., 2012. Structural plasticity of the left caudate in bimodal bilinguals. Cortex 48 (9), 1197–1206. http://dx.doi.org/10.1016/j.cortex.2011.05.022.