Exploring Individual Differences in Grey Matter Relating to Language Learning in Adulthood

Rebecca Pearse

McGill ID: 260777423

Integrated Program in Neuroscience (IPN), Faculty of Medicine

McGill University, Montreal, QC. Canada

Date of Final Thesis Submission: November 28th, 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science (MSc) 2019

Co-Supervisors: Dr Denise Klein and Dr Shari Baum

Table of Contents

Appendix: Testing Materials

Supplementary Materials

I. Abstract

There is considerable individual variation in successful language learning in adulthood. Individual differences in natural 'aptitude' for language learning may contribute to this variation and partly manifest in pre-training brain structure characteristics or 'biomarkers'. In this study, we used anatomical Magnetic Resonance Imaging and Voxel-Based Morphometry to explore whether a priori differences in grey matter volume (GMV) correlated with how successfully 18 adults (consisting of a group of 10 English-L1 speakers and a group of 8 Mandarin-L1, English-L2 speakers) learned French over 12 weeks of intensive training. We also explored changes in GMV after training associated with neuroplasticity. The focus of our behavioral measures was to quantify improvement in French articulation more objectively than in previous studies. We used speech production tasks including oral repetition, reading, and free speech, and precise acoustic analyses of changes in Voice Onset Time (VOT) and vowel formants, which can index speech nativeness.

Our main results showed: (1) improvement in French articulation in both L1 groups, including in the acoustic measures of VOT and /i/ vowel production; (2) no neuroplasticity effects; and (3) pre-training GMV biomarkers for French articulation (indexed by improvement on acoustic measures, articulation rate, and oral repetition) in brain regions linked to implicit motor learning (the bilateral cerebellum and the left basal ganglia) and phonological processing (the left caudate, the right inferior parietal lobe, and the right cerebellum). Unique biomarkers likely linked to the cognitive demands of specific tasks requiring reading and oral repetition also emerged.

These findings help to elucidate the structural neural basis of language aptitude and clarify sources of variability for articulatory learning success in adulthood. Some of this variability is due to a priori brain structure in regions underlying motor, phonological, and perceptual abilities.

II. Résumé

Il y a une variabilité considérable dans l'apprentissage fructueux des langues à l'âge adulte. Les différences individuelles de l'aptitude naturelle d'apprendre les langues peuvent contribuer à cette variation et en partie peuvent manifester d'une préformation des caractéristiques de la structure cérébrale ou des 'biomarqueurs'.Dans cette étude, nous avons utilisé Voxel-Based Morphometry (VBM) pour explorer si a priori les différences du volume de matière grise (GMV) a correlé avec le succès de 18 adultes (composés d'un groupe de 10 anglais monolingues et un groupe de 8 mandarin-anglais bilinguals) qui ont appris le français au cours de 12 semaines de formation intensive. Aussi nous avons exploré des changements de GMV après la formation associée avec la neuroplasticité. L'objet principal de nos mesures comportementales c'était de quantifier l'amélioration de l'articulation française plus objectivement que dans les études précédentes, utilisant les tâches de production de la parole, y compris la répétition orale, la lecture, et le discours libre, et des analyses acoustiques precises des changements du Voice Onset Time (VOT) et de formants des voyelles, qui peuvent indexer la parole autochone.

Nos résultats principaux ont montré: (1) une amélioration d'articulation en français dans les deux groupes, y compris dans les mesures acoustiques du VOT et /i/ production voyelle; (2) pas d'effets neuroplastiques; et (3) une préformation de biomarqueurs GMV pour l'amélioration d'articulation française (par l'acoustique, le taux d'articulation et les mesures de répétition orale) dans les régions cérébrales liées à l'apprentissage motrice implicite (le cervelet bilatéral et les noyaux gris centraux de la côté gauche) et le traitement phonologique (le caudé gauche, le lobe droit pariétal inférieur (IBL), et le cervelet droit). Des biomarqueurs uniques, peut-être liés aux exigences cognitives des tâches spécifiques, réquerant ou la lecture ou la répétition orale ont emergé aussi.

Ces conclusions aident à élucider la base structural neurale de l'aptitude linguistique et clarifient les sources de variabilité pour la réussite de l'apprentissage articulatoire à l'âge adulte. Une proportion de cette variabilité est à cause d'une structure cérébrale a priori dans les régions sous-jacentes des capacités motrices, phonologiques, et perceptuelles.

III. Acknowledgements

I would like to thank many people who have supported me during my MSc, without whom it would not have been possible.

I would chiefly like to thank my co-supervisors Dr Denise Klein and Dr Shari Baum for their committed advice and assistance. I am very grateful for all the time and insight that they have offered. I additionally thank my Advisory Committee members Dr Michael Petrides and Boris Bernhardt for their direction and guidance.

In terms of contributions from Klein and Baum lab members, I would like to sincerely thank my fellow students Kaija Sander and Tess Fairchild for their help, and Dr Jen-Kai Chen, Dr Annie Gilbert, and Dr Haruko Saito for their patience and expertise.

Finally, I would like to thank my amazing family and friends, both in Montreal and at home in England who have supported me throughout my time at University.

IV. Preface and Contribution of Authors

My co-supervisors Shari Baum (PhD)[ab] and Denise Klein (PhD) [bc] contributed to the design of the present study, the acquisition, analysis and interpretation of data, and editorial help in the preparation of this thesis.

I also received assistance from various students, research assistants, and members of the Klein and Baum labs. Specifically, Kaija Sander[d] was fundamental to the study design, participant testing, and data collection, and Tess Fairchild[d] contributed substantially to the vowel formant analysis. I also received critical guidance from Dr Jen-Kai Chen[c] in our neuroimaging analyses, and from Dr Annie Gilbert[a] and Dr Haruka Saito[a] in the acoustic analysis of speech samples.

All material appearing in this document was written by Rebecca Pearse. Figures were designed and created by Rebecca Pearse.

The research was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant and donation from the Blema and Arnold Steinberg Family Foundation to the Klein lab and by funds from the Fonds de Recherche Société et Culture Quebec (FRQSC) and the Social Sciences and Humanities Research Council (SSHRC) to the Baum lab.

I also received personal funding from the Canadian Rhodes Scholarship Foundation.

---

[a] School of Communication Sciences and Disorders, McGill University, Montreal, Quebec, Canada
[b] Centre for Research on Brain, Language, and Music (CRBLM), Montreal, Quebec, Canada
[c] Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
[d] Integrated Program in Neuroscience, McGill University, Montreal, Quebec, Canada

1. Introduction

In our increasingly globalised world multilingualism is becoming more prevalent, with more than half the global population actively learning or speaking a second (L2) or third language (L3) (Grosjean and Li, 2013). As opposed to learning multiple languages from birth, many individuals learn a new language in adulthood. For example, upon immigration to a new country. Such learning can afford profound socioeconomic opportunities, so is important to understand and facilitate. However, language acquisition in adulthood compared to childhood is a notoriously difficult task (Hull and Vaid, 2007), with significant individual variability in success (Dornyei, 2009).

Factors that affect adult language learning can be considered as either external to the learner, such as the learning environment, or internal to the learner, such as pre-existing language background or cognitive abilities (Wong and Ettlinger, 2011). A crucial difference in language background is the typological distance between the individual's existing language(s) and the target language. Together, the internal factors can be conceptualised as representing 'language aptitude' or 'neural preparedness' for second language learning. This aptitude may partly manifest in structural biomarkers that relate to subsequent learning success. For example, a priori variation in the morphology of a specific brain region may correlate with variation in subsequent learning due to differences in pre-existing neural efficiency or capacity (Golestani and Price, 2011).

In the present study, we used anatomical images from Magnetic Resonance Imaging (MRI) to conduct Voxel Based Morphometry (VBM) analyses in participants before and after 12 weeks of intensive French training. We explored grey matter volume (GMV) neuroplasticity effects after training, and whether pre-training regional differences in GMV related to subsequent French behavioral improvement across different tasks measuring proficiency of articulation.

In recognition of how language background can affect learning, we incorporated a group comparison into our analysis. One group consisted of 10 English-L1 monolinguals who were learning French as an L2, and the other consisted of 8 Mandarin-L1 speakers who also spoke English to varying degrees of proficiency and were learning French as an L3. The groups were matched on demographic factors, intelligence, cognitive control, working memory and non-native phonological discrimination. The comparison was designed to assess the impact of typological distance between languages on our measures of French behavioural improvement, neuroplasticity and biomarkers, since English and French are more similar than Mandarin and French.

Language processing is multidimensional, involving conceptual, perceptual and motor abilities. In the present study, we focused on measuring French articulation: the production of clear and nativelike speech sounds. Previous research in language learning has used subjective accent ratings instead of precise linguistic analysis to assess nativelike articulation. Measuring acoustic aspects of speech enables a more quantitative and fine-grained approach to analysing improvement in speech production and to exploring the subtle brain differences that may underlie articulatory learning. Therefore, we measured changes in our participants' Voice Onset Time (VOT) and vowel formant frequencies after French learning, which can index the nativeness of speech.

Overall, the present study helps to elucidate the neural structure that may partially predict variability in adult language learning. This outcome thus enables linking of characteristics in brain structure and function that promote articulatory learning, such as implicit motor coordination and phonological working memory, which may be modifiable by targeted training regimes. Establishing the conditions that best facilitate language learning has further applications to both educational and clinical practice, such as for designing individualised pedagogical programmes and in post-stroke speech rehabilitation where adults must relearn language.

2 Literature Review

In this section I summarise the previous research that led to the formulation of the current research questions and hypotheses. I first discuss language learning in adulthood, followed by the literature on neuroplasticity after language training and biomarkers for predicting language learning success. Next, I present the acoustic measures of speech analysis that we used in the current study and the technique of VBM. Finally, to develop our anatomical hypotheses, I discuss models of the neurobiology of speech production.

2.1 Language Learning in Adulthood

In childhood, new languages are acquired relatively effortlessly. By contrast, in adulthood, learning a new language after a putative 'sensitive period' is a hugely challenging task (Lenneberg, 1967; Long, 1990; Birdsong, 1999). For example, it is well documented that learning non-native speech categories as an adult is profoundly difficult (Best, 1995; Flege, 1995). Consistent with this difficulty, there is significant individual variability in outcome after language learning in adulthood (Golestani and Zatorre, 2009). Part of the reason for this variability is the complex interaction between numerous internal and external factors that predict learning success (Wong and Ettlinger, 2011), with internal factors pertaining to characteristics of the individual learner and external factors pertaining to the learning environment, such as the type and quality of input (Perrachione, 2011).

The internal factors that influence the ease of learning and therefore help to predict successful adult language learning can heuristically be considered as representing 'language aptitude'. These factors include general cognitive and learning factors and native language background, which will be addressed in turn.

2.1.1 *Cognitive and Learning Factors*

Language learning may emphasise different cognitive systems at different ages (Krashen, 1981). For example, children typically learn grammar implicitly (i.e. subconsciously), whereas adults rely largely on explicit (i.e. deliberate) learning especially in a classroom setting (Ullman, 2016). This hypothesis has typically been related specifically to grammar acquisition. Nevertheless, elements of speech production may also rely on explicit learning processes. For example, outcomes demanding phonological working memory such as sentence repetition. Explicit learning, such as memorisation of grammar rules, tends to demand domain-general declarative memory and reasoning processes (Frankish, 2010). Consistently, there is evidence linking domain-general cognitive abilities to variation in adult L2 learning. Notably, working memory capacity has been implicated as a core element in L2 aptitude (Robinson, 2005) and a predictor of overall language proficiency (Van der Noort, Bosch and Hugdahl, 2006).

However, some aspects of language are difficult to learn explicitly. In particular, acquiring the correct pronunciation and articulation of a new language may instead demand implicit motor and sensorimotor learning, which are not necessarily under conscious control. As such, researchers have also promoted the importance of procedural learning in predicting successful language acquisition (Linck et al., 2014), and the superiority of “reflexive” (implicit) learning systems over “reflective” (explicit) learning systems for perceptual speech category learning in adulthood (Yi, Maddox, Mumford, and Chandrasekaran, 2014). Therefore, the neural processes that underlie language aptitude include core, general abilities that partially depend on what aspects of speech and language are being tested. For example, with a focus on measuring articulation, existing motor coordination and phonological learning are likely to be relevant to the acquisition and adaptation of new speech sounds.

2.1.2 Native Language Background

Languages can differ dramatically in typology and similarity (Haspelmath, 2001). For example, English and French are more typologically similar than Mandarin and French in terms of their phonology, grammatical morphology and writing systems. For example, auxiliaries, articles, and verb inflections convey information on tense and meaning in English and French (Barac and Bialystok, 2012). In contrast, Mandarin is an uninflected, tonal language, whereby word meanings are distinguished by pitch patterns and tense is conveyed by context, adverbials, and word order (Yip, 2002).

A large typological mismatch can limit and slow the rate of learning in the L2 and the L3 when comparing the target language to the native language (Cenoz, 2001; Swain et al., 1990). This effect is due in part to increased difficulty in language switching (Yang et al., 2018). Moreover, reduced phonetic similarity of speech categories, such as between English and Mandarin, can lead to difficulty in phonological learning as learners usually become perceptually 'tuned' to their native speech categories at a young age (Flege, 1995; Kuhl, 2004). Consistently, although there is a claim that bilinguals outperform monolinguals when learning a new language (Abu-rabia and Sanitsky, 2010; Cenoz, 2003; Bartolotti and Marian, 2012), purportedly mediated by mechanisms such as more efficient phonological working memory (Kaushanskaya and Marian, 2009), bilingualism per se may not guarantee learning success over monolinguals if the relevant languages are phonetically dissimilar (Antoniou and Liang, 2015).

Overall, language background is an important factor to consider in learning studies. Language experience can heavily influence the ease of learning new grammar and phonological patterns, and therefore contribute to inherent aptitude for learning a new language.

2.2 Neuroplasticity

In cognitive neuroscience, the term neuroplasticity describes changes in neural structure or function over time that affect behaviour and are related to experience or training (Herholz and Zatorre, 2012; Maguire et al., 2000). An example of such experience is language acquisition. Factors that may affect the rate and scope of neuroplasticity in this context are age, proficiency level, language specific characteristics, and individual differences (Li, Legault and Litcofsky, 2014).

In terms of age, although some researchers have attested that language plasticity occurs during development and decreases across the life-span (Lenneberg, 1967; Newport, 2001; Kennedy and Raz, 2009), late language learning may actually be an ideal testbed to examine effects of neuroplasticity. When acquisition occurs past childhood it is largely dependent on explicit learning mechanisms (Ullman, 2016), and might lead to greater neural changes during the most cognitively demanding learning phases (Xiang et al., 2015). This stance means that brain plasticity is still possible during adulthood, albeit in a more restricted and specific manner (Bavelier and Levi, 2010). This conclusion has been supported by many studies that have aimed to record the effects of adult language learning on the brain, which will be discussed below.

2.2.1 Language Training Studies

Most studies of neuroplasticity have explored bilinguals with long-term, sometimes lifelong, L2 experience. However, many individuals attempt to learn an L2 or L3 during adulthood over a shorter period of time. For example, in an intensive language course upon immigration to a foreign country. This experience can be an intense cognitive training regime and induce both functional and structural neuroplasticity (Costa and Sebastian-Galles, 2014).

Considering studies using a learning schedule of less than 6 months, Stein et al. (2012) investigated GMV changes in a group of college-age English learners of German after 5 months of immersive training. The authors reported an increase in GMV in the left inferior frontal gyrus (IFG) and anterior temporal lobe, which are areas implicated in lexical access and semantic integration. However, the study lacked a control group, meaning it is impossible to rule out confounding factors like mixed L1 language background or motivation. Nevertheless, other studies have reported convincing plasticity effects over similar time scales. For example, Barbeau et al. (2016) used functional MRI to report an increase in inferior parietal lobe (IPL) activation during L2 sentence reading after 12 weeks of French learning by English speakers.

Studies using slightly longer training schedules have indicated more consistent neuroplasticity effects. For example, Schlegel et al. (2012) explored longitudinal differences in white matter connectivity using Diffusion Tensor Imaging (DTI) in college-age students undertaking an intensive Chinese course (7.5 hours a week for 9 months). Results showed increased fractional anisotropy (FA), particularly in the left caudate nucleus, which was directly associated with behavioral improvement. This finding is in line with other studies implicating the caudate in language control (Green and Abutalebi, 2008; Branzi et al., 2015).

These results imply that articulation learning may also induce structural neuroplasticity in similar regions to those previously reported, given the importance of the left caudate in language control, and the likely role of phonological processing regions such as the IPL in learning new articulation patterns. However, most studies investigating neuroplasticity after language training focus on reporting the structural or functional consequences of such training, and do not offer insight into what pre-training neural factors may actually constitute ‘biomarkers’ for predicting subsequent learning.

2.3 Biomarkers

In cognitive neuroscience, the term biomarker refers to structural or functional characteristics that predict positive or negative behavioural or neural effects. The rationale for investigating neural biomarkers in language learning is based on the observation that some adults have difficulties at every step of language learning, while others can grasp the skill in less time and with more success, given the same learning conditions. For example, adults with very similar language backgrounds can differ considerably in their ability to perceive non-native speech sounds following phonetic training (Golestani and Zatorre, 2004). Therefore, individuals may vary in their 'aptitude' or 'neural preparedness' for language acquisition, which may manifest in pre-training neural structural differences that contribute to the variability in success (Dornyei, 2006). This idea has been the basis for various studies aiming to elucidate the neural basis of language aptitude, which will be summarised below.

2.3.1 Language Training Studies

A number of recent studies have explored if structural or functional neural biomarkers can predict how successfully an adult learns a new language. For example, Chai et al. (2016) used resting state fMRI to determine if individual differences in functional connectivity related to French acquisition in English speakers completing a 12-week intensive course. The authors focussed on lexical retrieval in spontaneous speech and reading speed as their measures of proficiency, and on the anterior insula and visual word form area (VWFA) as regions of interest, in line with previous findings (Veroude et al., 2010). Results showed that pre-training connectivity between the left anterior insula and dorsal anterior cingulate cortex (ACC) and between the left anterior insula and left posterior superior temporal gyrus (STG) correlated positively with improvement in L2 lexical retrieval. Second, connectivity between the VWFA and left mid- STG correlated positively with improvement in L2 reading speed. Results were interpreted by Chai et al. as indicating that a priori individual variability in connectivity of language-relevant networks influences subsequent variability in aptitude for L2 learning. Moreover, variability in connectivity of different networks can predict improvement in different tasks, meaning aptitude for language learning may depend on the specific behavioral measure.

Studies have also investigated structural predictors for short-term auditory and auditory-motor learning. These studies are relevant to considering potential structural predictors for more long-term, real-world articulatory learning since articulation is partly based on successful auditory-motor processing. For example, in speech perception, faster versus slower learners of non-native phonemic speech contrasts showed higher pre-existing white matter density (WMD) in left Heschl's Gyrus (Golestani et al., 2006) and bilateral parietal regions (Golestani, Paus and Zatorre, 2002). Meanwhile, in speech production, Golestani and Pallier (2006) reported that WMD but not grey matter density (GMD) differences in the left insula and bilateral IPL regions were associated with accurate pronunciation of non-native phonemes.

Therefore, these results indicate that pre-existing structural variability in acoustic and auditory-motor processing areas, particularly the IPL, may serve as powerful predictors of future phonological learning. Nevertheless, it is also important to remember that biomarkers for learning success may depend on the specific typology and similarity of the native and target languages in question, so may not serve as universal predictors of success across language combinations (Qi et al., 2015). Moreover, language learning studies in this context have mainly used lexical measures of language improvement, such as lexical retrieval, as opposed to perhaps more detailed, sub-lexical indicators of nativelike speech production. Good candidates for such indicators include acoustic speech measures.

2.4 Acoustic Speech Measures

Previous studies of language training have used speech samples to extract indices of speech and language proficiency. For example, Berken et al. (2016) used subjective ratings by native speakers to measure accent in his cohort, and Golestani et al. (2006) used native speakers to judge phoneme production “goodness”. However, there was no objective investigation into acoustic features of speech which may be indicative of the nativeness of articulation, such as VOT and vowel production. Acoustic speech measures can offer a more quantitative and fine-grained approach to measuring accurate pronunciation, and therefore allow for more detailed investigation of potential brain differences that are relevant to articulatory learning.

2.4.1 Voice Onset Time (VOT)

VOT is an acoustic feature of speech related to the production of stop consonants (i.e. plosives) in which the vocal tract is momentarily blocked so that airflow ceases, such as the /b/ sound at the beginning of the word “bat”. Plosives can either be voiced (e.g. /b/) or voiceless (e.g. /p/), which refers to whether the sound involves vibration of the vocal cords or not (Klatt, 1975). Plosives also differ by Place of Articulation (PoA), which refers to the point of contact where the air flow is obstructed in the vocal tract. For example, the sounds /p/ and /b/ are bi-labial because the sounds are produced by using both lips together (Stevens and Blumstein, 1978). VOT is defined by the length of time between the release of the stop consonant and the onset of voicing, which can be measured on the waveform of a speech signal.

VOT may be useful to index the ‘French-ness’ of speech because the VOT for stop consonants differs between English and French, with different VOT categories differentiating voiced and voiceless stops. Specifically, English speakers produce long-lag VOTs for voiceless stops and short-lag VOTs for voiced stops (Docherty, 1992), whereas French speakers' voiceless stops are in the short-lag VOT range, and their voiced stops are pre-voiced (Ryalls and Larouche, 1992). These findings are supported by studies showing that French-English bilinguals often exhibit 'intermediate' VOT patterns (Flege, 1987). Therefore, VOT provides an acoustic index of the nativeness of speech production.

2.4.2 Vowels

Vowel production may also provide a useful measure of speech nativeness. Vowels are produced as the vocal cords vibrate, creating a periodic, harmonically complex sound with a pitch that is associated with the fundamental frequency (F0) of the vocal fold vibration. As the sound passes through the vocal tract, the shapes of the anatomical structures reinforce the acoustic energy in certain frequency bands. The acoustic energies at these resonance frequencies are called formants, which can be observed on a speech waveform in the form of energy bands and changed by altering the shape of the vocal tract (Hillenbrand, 2006). The first 2 formants (F1 and F2) play a significant role in vowel identity, with F1 inversely related to vowel (tongue) height and F2 to vowel back-ness (Kent and Read, 1992). Vowel formant frequencies differ between English and French in some cases and can be affected by language experience. For example, Flege et al. (1997) reported effects of experience on non-native speakers' production of English vowels, suggesting that people can acquire new vowel categories as they improve in the speech production of a target language.

Therefore, vowel production also provides an acoustic index of the nativeness of speech production and allows investigation of subtle brain differences potentially underlying articulatory learning. Relevant brain differences may include structural variation in regional grey and white matter, which can be explored with the technique of Voxel Based Morphometry.

2.5 Voxel Based Morphometry (VBM)

VBM is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue. Through a voxel-wise comparison of multiple structural magnetic resonance images, it can indicate regions where grey or white matter differ significantly between groups. VBM can also allow extraction of GMV or GMD in specific brain voxels across individuals, to later correlate with behavioral measures. A value of VBM is that it enables comprehensive measurement of grey matter differences not just in cortical structures, but throughout the entire brain, which may result from structural reorganisation induced by experience, or genetic predisposition (Mechelli et al., 2005). Although researchers can vary in their interpretation of VBM, in general higher GMD and GMV is thought to reflect positive qualities such as higher processing capacity or efficiency (Elmer et al. 2014).

VBM has been widely used in investigations of brain regions implicated in language processing. For example, Golestani et al. (2002) used exploratory VBM to investigate structural predictors for the rate of learning non-native phonetic speech contrasts, although over a single training session as opposed to real language learning. Since VBM is a whole-brain measure, it allows investigating potential GMV biomarkers for language learning in both cortical and subcortical structures, informed by predictions from models of speech production, which will be discussed below.

2.6 Speech and Language Brain Regions

2.6.1 Speech Production

Speech production can be measured by many different tasks, such as free speech, sentence repetition, and reading speed. These tasks require partially distinct neural systems, for example based on differences in cognitive demands, and therefore may yield different neural biomarkers for learning improvement. For example, differences in brain structure or function in the left Visual Word Form Area (VWFA) (Dehaene et al., 2010) and the left IPL (Barbeau et al. 2016) may predict improvement in reading speed. Meanwhile, differences in regions purportedly involved in auditory sentence processing, specifically temporal regions for semantic processing and identification, and frontal regions for building syntactic and semantic relations (Friederici, 2002), might be especially important for predicting improvement in sentence repetition.

Nevertheless, different tasks measuring speech production still have the common requirement of accurate speech articulation, so may also show common biomarkers in similar brain regions. Specifically, structure in brain regions related to phonological processing and implicit motor learning might predict improvement across measures of articulation, including lexical measures of fluency such as lexical retrieval and sub-lexical measures such as acoustic speech factors. Learning accurate and fluent articulation in a foreign language involves a host of neural processes, but critically depends on discriminating phonological patterns and specific motor coordination of articulators, which will be discussed in turn.

Regarding phonological processing, brain regions in the temporo-parietal cortex are critical (Tremblay and Dick, 2016). For example, greater WMD in the IPL (including the supramarginal gyrus (SMG) and the angular gyrus (AG)) was reported in faster learners of non-native phoneme perception and production (Golestani et al., 2002; Golestani et al., 2006). The IPL is an integral part of the phonological network (Hickok and Poeppel, 2007) and has been conceptualised as a temporary buffer in phonological working memory (Koelsch et al. 2009).

Additionally, subcortical regions may contribute to phonological processing. For example, higher GMD in the bilateral caudate head was associated with better performance in a phonemic fluency task in the L2 of high-proficiency bilinguals (Grogan et al. 2009). The caudate is also activated in feedback driven phonological learning tasks, and in detecting phonological anomalies (Tettamanti et al. 2005b). Finally, the right inferior cerebellum (specifically lobule VIIb) has been implicated in phonological working memory circuits alongside the IPL (Chen and Desmond, 2005). Overall, structural biomarkers in these areas underlying phonological processing may predict articulation improvement in a new language due to their role in discriminating and learning new speech-sound patterns.

Regarding motor learning, brain regions can broadly be divided into cortical regions for explicit processing and subcortical regions for implicit processing, which may behave competitively (Kantak et al., 2012). Cortical regions including the dorsolateral prefrontal cortex (dlPFC), the bilateral premotor cortex (PMC) and the supplementary motor area (SMA) are thought to develop explicit knowledge of motor sequences (Honda et al., 1998). In contrast, subcortically, the basal ganglia are typically thought to sub-serve procedural motor and sequence learning (Ullman, 2004), as well as speech motor control. For example, imaging studies have reported left caudate activity for tasks demanding the selection of competing verbal responses and language switching (Crinion et al., 2006; Zou et al., 2012), and the left putamen has similarly been implicated in articulatory control during L2 speaking (Klein et al. 1994). Otherwise, the bilateral cerebellum is important for sensorimotor integration and the motor adaptation aspect of implicit motor learning (Doyon et al. 2003). Specifically, the cerebellum is speculated to drive corrective motor commands through generating sensory error during movement (Grafton et al. 2008). Overall, structural biomarkers in these areas underlying implicit motor coordination may predict articulation improvement in a new language due to their potential role in motor control of articulators during speech.

The division between implicit and explicit motor learning reflects previous work into dual-learning systems during speech category learning in adults (Chandrasekaran, Koslov and Maddox, 2014). Specifically, Yi et al., (2014) used fMRI to report that adults optimally learn speech categories via cortico-striatal circuitry involved in slow, "reflexive" (procedural-based) learning (Seger and Miller, 2010) in which stimuli are implicitly associated with category responses. The reflexive system critically included cortico-striatal loops, as well as auditory regions including the mid temporal gyrus (MTG), and the IPL. Implicit learning contrasted an explicit "reflective" system, that was fast, feedback-dependent and encoded within sensory areas connected to the dlPFC, which is thought to generate verbalised rules which are then retained or discarded by the ACC (Ashby and Maddox, 2005).

This division between reflective and reflexive speech category learning could be applied to acoustic indicators of French articulation, since VOT and vowel production are not typically under conscious motor control. In addition, the multidimensional, highly redundant, and variable nature of acoustic cues in speech categories across talkers means that creating rules for such a large dimensional space may be sup-optimal (Holt and Lotto, 2010). Consistently, studies such as Soros et al. (2006) have asserted key roles of subcortical motor areas including the basal ganglia, thalamus and cerebellum in vowel production. Therefore, structure in brain regions important for reflexive learning may predict acoustic indicators of French articulation ability.

2.6.2 Effect of Language Background

Bilingual experience can have a marked and specific impact on the organisation of neural systems involved in language control, articulation, and sensory processing (Neville and Bavelier, 1998), due in part to the need to switch between and manage different lexicons and articulatory repertoires.

Abutalebi and Green (2007) delineate an integrated neural network important for control during bilingual speech production, including cortical regions such as the ACC, the IFG, and the IPL, and subcortical regions such as the caudate nucleus (Grogan et al., 2012) and putamen (Abutalebi et al., 2013). Of particular note, Abutalebi et al. (2011) argue that bilingualism 'tunes' the ACC for efficiency in domain-general conflict monitoring and is associated with increased GMD in the ACC (Abutalebi et al., 2015). As such, structural neural differences may occur between monolingual and bilingual learners in these brain regions thought to be especially influenced by previous language experience.

Finally, studies have corroborated the putamen as important for L2 articulation. For example, Berken et al. (2016) compared simultaneous bilinguals (who learnt 2 languages from birth) and highly proficient sequential bilinguals (who learnt 2 languages sequentially), who differed only in the degree of their native accent. Between groups, the simultaneous bilinguals had increased GMD in regions including the left putamen and left posterior insula, while, the sequential bilinguals with better accents also showed greater GMD in the left putamen, pinpointing this region as relating specifically to accent and articulation.

To conclude, structural biomarkers for articulation learning success in adulthood may occur in many different brain regions important for speech production. Unique biomarkers may emerge according to the specific cognitive requirements of different speech production tasks, as well as common biomarkers across tasks relating to the general requirements of learning the nativelike articulation of a new language. Of particular importance may be brain regions involved in fine, implicit motor coordination of articulators and brain regions important in phonological learning and working memory. Finally, structural differences in biomarkers may occur in brain regions thought to differ between monolinguals and bilinguals.

3 The Present Study; Research Objectives and Predictions

1) Acoustic Analysis

We aimed to use VOT and vowel formant analysis to index improvement in French articulation in adult learners in a more quantitative manner than native accent ratings. Following 12 weeks of intensive training, we predicted participants to show learning effects by producing vowel formant frequencies closer to French typical values and shorter average VOT durations.

2) Neuroplasticity

We aimed to explore GMV neuroplasticity and predicted plastic effects in speech and language-related neural circuits after 12 weeks of intensive French training in adult participants.

3) Biomarkers

We aimed to use whole-brain VBM analyses to detect structural neural biomarkers for French articulation improvement in adulthood. We predicted that a-priori GMV differences in regions important for motor coordination and phonological processing would correlate with behavioral improvement across different articulatory measures (e.g. acoustic measures, oral repetition, reading speed), and that unique biomarkers may also be evident according to the specific sensory or cognitive requirements of different tasks (e.g. reading or oral repetition).

4) Language Background

We aimed to compare the effect of similarity between the L1 (English or Mandarin) and the target language (French) on behavioral improvement and GMV language biomarkers. We predicted that the English-L1 group may show more French articulation improvement, and that there may be group biomarker differences between English-L1 and Mandarin-L1 learners.

4 Methodology

4.1 Participants

We tested 18 participants (12 females and 6 males) between the ages of 17 and 32 (*Mean*: 21.07, *SD:* 3 years). There were 10 monolingual American-English native speakers (the English-L1 group) (*Mean age*: 20, *SD*: 2.3 years), and 8 speakers with Mandarin as their native language and English as their L2 (the Mandarin-L1 group) (*Mean age*: 22, *SD*: 4.7 years).

Participants were recruited from a McGill Language Centre course for beginner learners of French. The course curriculum included both literacy and conversational training. Students were provided with a description of the study and recruited if they agreed to be contacted and met our criteria. The study was approved by the Research Ethics Board of the Montreal Neurological Institute (MNI), and participants gave written informed consent.

Participants were all right-handed, as assessed by the Handedness Inventory (Briggs and Nebes, 1975). Participants exhibited a range of musical abilities, but there were no professional musicians included given the link between musical training and brain organisation (Gaser and Schlaug, 2003). All participants were attending McGill University at the time of study. Exclusion criteria included all known hearing, vision or reading impairment, history of brain trauma or neurological disorder, and conditions incompatible with MRI such as metal implants. These criteria were assessed by the Health Information questionnaire (see **Appendix I**.).

Participants were given an in-depth language questionnaire (Language Experience and Proficiency Questionnaire (LEAP-Q), Marian et al., 2007) which measured number of years of language experience, accent ratings, and language education, among other variables (See **Appendix I**).

The English-L1 group included American or Canadian students from outside of Quebec who may have had some previous exposure to French but considered themselves beginner learners. Some of these participants had some knowledge of other languages (usually Spanish) but had a mean % use of English in their daily conversations of 90%. The Mandarin-L1 group reported more balanced use of their 2 languages (44% conversation use for English) and a mean English Age of Acquisition (AoA) of 8 years (see **Table 1**).

Table 1. Participant Demographics

Group	N	Gender (male/female)	Chronological Age (years)	English Age of Acquisition (years)	English use % conversations
English	10	4/6	20 *(2.3)*	From Birth	90 *(8.1)*
Mandarin	8	2/6	22 *(4.7)*	8 *(2.5)*	44 *(14.7)*
t Statistic	_	_	*-1.19*	_	*8.46*
P value	_	_	*.253*	_	*< .001**

Mean *(SD)*

The Mandarin-L1 and English-L1 groups were matched on intelligence. There were no group differences in performance in the Matrix Reasoning subtest, the Letter-Number Sequencing subtest, and the Digit Span subtest of the WAIS-IV (Wechsler Abbreviated Scale of Intelligence; Weschler, 2008) (See **Table 2**).

The groups were also matched on cognitive control. There were no group differences in performance on a modified version of the Simon Task (Simon and Rudell, 1967) used to measure the stimulus-response compatibility effect, as per the method of Kousaie et al. (2017). Outcome measures were the Simon effect, interference suppression, and response inhibition. (See **Table 2**). See the **Supplementary Materials** (section **I**) for task measurement and calculation details.

Table 2. Cognitive and Intelligence Measures

	English-L1	Mandarin-L1	*t Statistic*	*P Value*
Matrix Reasoning (/26)	22.5 *(1.12)*	22 *(1.58)*	*0.79*	*.443*
Letter-Number Sequencing (/30)	20.1 *(1.3)*	19.8 *(1.3)*	*0.49*	*.633*
Digit Span: Forward (/16)	11.7 *(1.70)*	10.25 *(1.91)*	*1.70*	*.108*
Digit Span: Backward (/16)	8.6 *(2.01)*	9.4 *(2.07)*	*0.83*	*.420*
Digit Span: Sequencing (/16)	8.3 *(1.06)*	8.1 *(2.30)*	*0.25*	*.809*
Simon Effect (msec)	17.3 *(20.1)*	8.8 *(17.9)*	*0.94*	*.364*
Interference Suppression	116.1 *(31.8)*	119.9 *(29.5)*	*-0.26*	*.798*
Response Inhibition	68.2 *(57.2)*	104.2 *(46.5)*	*-1.44*	*.170*

Mean *(SD)*. Maximum possible scores where applicable in column 1 parentheses.

4.2 Procedure and Analysis

Participants were tested behaviourally and given structural Magnetic Resonance Imaging (MRI) scans pre and post participation in an intensive beginner French language program. French training totalled 80 hours of classes over 12 weeks. The participants were also tested behaviourally after 4 weeks of classes, but this thesis focuses only on the pre versus post (12 week) training improvement. Behavioural testing included speech and language assessment in both French and English. The French measurements were designed to indicate improvement induced by the French training, while the English measurements were designed to serve as a control for test-retest effects, and to explore potential language transfer effects in the Mandarin-L1 group. The participants underwent the same behavioural testing session at both pre and post training. All statistical analyses were completed with IBM SPSS Statistics 24.

4.2.1 Behavioural Session

4.2.1.1 Speech and Language Assessment

To assess language proficiency and improvement, participants completed 5 speech and language tasks in both English and French at both pre and post training. The English version of the task always preceded the French version. All speech samples were digitally recorded at a distance of 50cm from the unidirectional, table-mounted microphone. The order and details of the tasks were as follows:

1. Story Reading Task

Participants read aloud one paragraph taken from the Story Learning and Memory test (Djordjevic et al., 2011) (duration approximately 2 minutes) and answered 16 questions about the content of the story (See **Appendix II**). The story was originally in English and was translated into French by a native speaker. This task allowed measurement of the number of words read per minute (Dehaene et al., 2010; Chai et al., 2016) and acoustic analysis of the participants' recorded speech production (described below). For this thesis, the exact content of the responses to the comprehension questions was not analysed.

2. Picture Description Task

Participants freely described a picture of a household scene in as much detail as possible for 2 minutes (Cookie Theft picture, Lightbulb picture, Boston Diagnostic Aphasia Exam, Kaplan and Goodglass, 1983) (See **Appendix III**). The pictures were different for French and English to avoid attempts at direct translation. The same picture was used for each language at each time-point to control for varying difficulty in the stimulus. Behavioral measures were extracted from the speech samples that indexed the rate, content, and accuracy of speech production (see **Table 3**). Measurement of pause duration and frequency was completed using the speech analysis software Praat (http://www.praat.org/: Boersma, Paul and Weenink, 2018).

French accent was also rated using the recordings of this task by 2 independent native Quebec-French speakers on a scale of 1 (very poor) to 7 (native-like), in a manner similar to Berken et al. (2016). The speech samples were presented in a random order and the raters were blind to the participant's L1 and the time of testing. Seven out of the 36 sound files were randomly selected and blindly reprocessed to check for consistency of judgment (within one point-score). The Cronbach's alpha score for consistency between the 2 independent ratings of accent improvement was low at 0.28, so the ratings were not used in the neuroimaging analyses.

Table 3. Measures extracted from the Picture Description Task

Measure	Description
Total Words	The total words spoken in 2 minutes, not including code-switching between languages, intrusions, and repetitions
Vocabulary Index	The sum of the number of correct and unique nouns, verbs, adjectives, prepositions, and determiners
Grammar Index	The proportion of complex and compound sentences relative to the total number of complex, compound, and simple sentences
Articulation Rate	The mean number of syllables produced per second
Mean number of Syllables per Run	The mean number of syllables produced consecutively without a 0.25 second pause
Filled Pauses	The number multiplied by the mean duration of filled pauses; defined as a sound with no lexical meaning of at least 0.2 seconds
Silent Pauses	The number multiplied by the mean duration of silent pauses; defined as an unfilled pause of at least 0.25 seconds

Table 3: the 7 measures extracted from the full 2-minute free speech samples of the Picture Description task, and how each measure was calculated.

3. Sentence Repetition Task

Participants completed the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals (4th edition) (Semel, Wiig and Secord, 2003), in the original English and translated into French by a native speaker. Twenty-four sentences were read aloud by the same tester to participants who repeated them immediately after hearing them (See **Appendix IV**). This task was scored by calculating the percentage of words correctly repeated (in grammar and pronunciation) for each of the sentences separately, then averaging the scores into a composite score of repetition accuracy. The recordings were also used for the acoustic analysis.

4. Verbal Fluency

Participants were asked to produce as many different and correct examples of a word beginning with a particular letter over 3 trials of 60 seconds each. The letters were F, A, and S for English and P, F, and L for French. These stimuli have been used extensively to examine verbal fluency in previous studies. This task was scored by the total words produced in 60 seconds across the 3 letters, not including errors and repetitions.

5. Hindi Discrimination Task

Participants were asked to discriminate non-native speech sounds in a modified, computerised version of the Hindi Phoneme Identification task (as per the method of Golestani and Zatorre, 2004). This task required perceptual discrimination of the Hindi dental-retroflex (/ta/-/ʈa/) stop consonants; a phonemic contrast that does not exist in English, French, or Mandarin. This task was scored out of a maximum of 600, which reflects the total number of trials.

4.2.1.2 Acoustic Analysis

Voice Onset Time

Voiced (/b/, /d/, /g/) and voiceless (/p/, /t/, /k/) stop consonants at the onset of a word or syllable were identified in the recorded Story Reading (**Appendix II**) and Sentence Repetition (**Appendix IV**) stimuli. Exclusion criteria included a stop consonant with an immediately following /l/ or /r/, which can change speech production. The plosives differed by pair in PoA (/p/ and /b/: bilabial; /t/ and /d/: alveolar; and /k/ and /g/: velar). In total, there were 120 potential voiced and 98 potential voiceless tokens for English, and 100 potential voiced and 236 potential voiceless tokens for French at both time-points.

VOT measurement consisted of visually locating the release of the stop consonant (the "burst" onset) and the onset of periodic energy in the waveform for each individual token using Praat. The duration between the 2 onsets reflected the VOT in milliseconds. If the onset of periodic energy preceded the onset of the burst, the VOT was negative (i.e. pre-voicing) (Baum et al., 1990). Measurements were cross-validated by 2 independent raters by confirming that a subset of VOT measurements were within 10% of each other. VOT values for each of the 6 plosives were analysed to produce individual measures of VOT change over time for each participant, reflecting the average change in VOT duration for French voiced and voiceless stop consonants between pre- and post-training.

Vowels

We conducted the vowel formant analysis by first labelling the vowels in the spectrograms of the Story Reading (**Appendix II**) and Sentence Repetition (**Appendix IV**) stimuli for each participant. The vowels studied for formant extraction were /i/, /u/, /ɑ/, /œ/, /e/, and /ɛ/ for French, with /ə/ instead of /œ/ for English since /œ/ does not exist in English. Due to varied task performance, we extracted n = 5 to 96 tokens of each vowel for each participant at each time-point (mean n = 22, SD = 20.15). For 3/18 participants the speech recordings were not clear enough to allow for vowel formant analysis and were excluded. The final analytic samples were 6 for the Mandarin-L1 group and 9 for the English-L1 group.

For each vowel, we extracted F1 and F2 frequencies from a 20ms window centered around the vowel midpoint (McFarland and Baum, 1995) using Praat. We explored the F1 vs F2 vowel space by plotting the formants produced by individuals pre- and post-training, compared to standard male or female French reference vowels (Kahn et al., 2011), and the average male or female English vowel production of our cohort across both native speaker groups. Separate male and female references were used because women typically have shorter vocal tracts than men and produce vowels with higher frequencies. This approach enabled calculating the Euclidean Distance (d) on the F2 vs F1 graph space between an individual's average vowel production (X_1 (F2), Y_1 (F1)) and the English or French reference vowels (X_2 (F2), Y_2 (F1)), and so the change in this Euclidean Distance over time (see **Figure 1**).

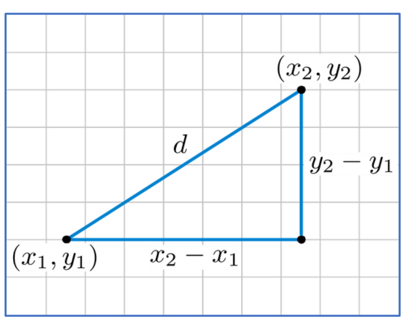


Figure 1. *Vowel Formant Analysis: d* represents the Euclidean Distance between an individual's vowel production (X_1 (F2), Y_1 (F1)) and the English or French reference vowels (X_2 (F2), Y_2 (F1)).

4.2.2 Structural Magnetic Resonance Imaging

4.2.2.1 Image Acquisition

All the imaging data were acquired at the McConnell Brain Imaging Centre at the MNI on a Siemens 3 Tesla MAGNETOM scanner. Participants underwent global 3-dimensional, T1-weighted scans at both pre-training and post-training. Images were obtained from a 3D Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence to acquire high resolution anatomical images (Matrix size: 256x256; Voxel size: 1 x 1 x 1 mm; Repetition time (TR): 2300ms; Echo time (TE): 2.96ms, Flip Angle: 9 degrees, FoV read: 256mm). Each scan lasted approximately 15 minutes.

4.2.2.2 Preprocessing

VBM analysis was carried out using the T1-weighted images. Data were preprocessed using the software platform CBRAIN and the CIVET 2.1.0 human brain-imaging pipeline. Standard spatial preprocessing steps were completed including (1) non-uniformity correction to remove variations in signal intensity related to radio-frequency inhomogeneity, used to ensure a more accurate tissue classification (Sled et al., 1998); (2) linear registration of corrected images into a stereotaxic space based on the standardised Montreal Neurological Institute 152 average template, used to normalise the images for between-subject differences in head position, brain size and shape, and so allow for comparison of anatomical data between subjects (Collins et al., 1994); (3) classification of brain tissue into white matter, grey matter, and cerebrospinal fluid using an automatic tissue-classification algorithm known as INSECT (Zijdenbos et al., 2002); (4) blurring of the binary grey matter map extracted from the classified image using a Gaussian smoothing kernel of 8 mm full width at half maximum, used to compensate for imperfect registration (Ashburner and Friston, 2001). This step converts binary data into continuous data, which is necessary for parametric statistical analysis including correlational analysis with other continuous variables (Golestani et al., 2002); and (5) modulation by Jacobian Determinants, which involves scaling by the amount of contraction or expansion that occurs through spatial normalisation. This step produces measures interpreted as grey matter volume, as compared to grey matter density for unmodulated images.

4.2.2.3 Second Level Analysis

Statistical analyses were carried out using MatLab and glim_image; an in-house software developed at the MNI. For all analyses, statistical threshold was established at $t = 4.3$, corresponding to $p < .05$ corrected significance using random field theory (RFT) (Worsley et al., 1996). RFT is appropriate to use in the current study since, due to the smoothing applied during pre-processing using a Gaussian kernel, each voxel is correlated with the neighbouring voxels so is not an independent observation. Therefore, the typical Bonferoni correction would be too conservative.

Neuroplasticity

Whole-brain, voxel-wise comparisons of GMV between pre-training and post-training were conducted for each of the L1 groups separately and across the 18 participants combined. Paired samples t-tests were conducted, and grey matter t-statistical maps were generated for each comparison to reflect the likelihood of longitudinal GMV changes.

More specific neuroplasticity analyses were also subsequently conducted according to the results of the biomarker analysis (described below). The GMV's at the specific voxel coordinates which showed biomarker effects were extracted at both pre and post-training for each participant, and Bonferroni corrected paired sample t-tests were conducted to compare the GMV values over time. This analysis was performed for a total of 12 specific voxel coordinates across all 18 participants.

Biomarkers

Whole-brain, voxel-wise regression analyses were used to investigate the relationship between GMV at pre-training and significant behavioral improvement for each of our L1 groups separately and across the 18 participants combined. GMV was the main predictor of behavioral improvement and age was a confounding factor. Separate regression analyses were performed for the behavioural tasks that showed significant improvement in French and no change in the equivalent English control task (see **Results**). These variables were (1) VOT; (2) /i/ Vowel Production; (3) Articulation Rate; (4) Sentence Repetition (accuracy); and (5) Words per Minute (reading).

T-statistic maps were produced corresponding to whether the regression slope between the GMV and behavioral variable of interest was significantly different from zero. Using the visualisation software Register (MacDonald, 2003), voxels with a t-value above statistical threshold were visually examined. The GMV values were extracted from the locations on the t-statistic maps with the highest t-values, and linear Pearson Correlation analyses were performed to calculate the correlation coefficient (*r*) between regional GMV and behavioural improvement on the task in question. Scatter plots were produced of these results.

The software NeuroSynth (http://neurosynth.org/) was used to help with the anatomical localisation of biomarkers. This platform uses the results of previously published articles to link voxel coordinates to different brain functions and areas, and therefore facilitate interpretation.

5 Results

5.1 Behavioral Results

5.1.1 Acoustic Analysis

Voice Onset Time

Average VOT durations in French and English speech for the 2 L1 groups separately at pre and post-training are shown in **Table 1** for voiced tokens (/b/, /d/, /g/) and **Table 2** for voiceless tokens (/p/, /t/, /k/) in the **Supplementary Materials** (section **II**).

French VOT values significantly changed over time. We ran a univariate 3-way ANOVA with PoA (3 levels: labial, dental, or velar), Voicing (2 levels: voiced or voiceless), and Time (2 levels: pre-training or post-training) as factors. Results showed significant main effects for Time (F (1, 204) = 14.96, p = <.001, ηp^2 = 0.068) and Voicing (F (1, 204) = 1936.97, p = <.001, ηp^2 = 0.905), but not for PoA. The 2-way interaction of Voicing*PoA was significant (F (2, 204) = 9.47, p = <.001, ηp^2 = 0.085), but not the remaining 2-way interactions nor the 3-way interaction. In contrast, English VOT values did not change significantly over time. We ran the same 3-way ANOVA comparing English VOT distributions at pre- and post-training and found no significant effect of Time.

Post-hoc paired sample t-tests (Bonferroni corrected) showed a significant difference between French pre-training and post-training for voiced tokens across the 2 L1 groups combined (t (17) = 3.15, p = .006). When the groups were separated, this difference was significant for Mandarin-L1 participants (t (7) = 2.44, p = .045), but not for English-L1 participants (t (9) = 2.12, p = .063) (see **Table 4**). For voiceless tokens, there was a significant difference between French pre-training and post-training across the groups combined (t (17) =

6.08, p = <.001). When the groups were separated, this difference was significant for the Mandarin-L1 participants (t (7) = 4.3, p = .004), and for the English-L1 participants (t (9) = 4.13, p = .003) (see **Table 4**). For each of these comparisons, the average VOT duration decreased over time, although the shift was not large.

Figures 2-3 show the change in French VOT per participant for the average duration of voiced and voiceless tokens respectively, split up by L1 group. It is possible to see that most individuals improve on VOT production as durations typically become shorter over time.

Table 4. French VOT Change

	Voiced			Voiceless			Combined
	Pre	Post	Change	Pre	Post	Change	Change
English-L1	20.6 *(2.5)*	19.6 *(2.9)*	-1.0 *(1.4)*	54.2 *(5.9)*	50.6 *(5.8)*	-3.6* *(2.6)*	-2.6 *(1.7)*
Mandarin-L1	21.5 *(3.1)*	19.3 *(3.4)*	-2.3* *(2.4)*	55.0 *(3.9)*	50.6 *(4.6)*	-4.4* *(2.7)*	-3.6 *(1.9)*
Average	21.0 *(3.9)*	19.4 *(3.6)*	-1.6* *(4.1)*	54.6 *(7.1)*	50.6 *(6.3)*	-3.9* *(7.2)*	-3.1 *(1.9)*

Table 4: the average combined voiced and voiceless VOT values (in msec) at pre and post training in French for the L1 groups separately. A negative Change value indicates improvement. Mean *(SD)*.

In summary, we found that: (1) for the L1 groups combined, both voiced and voiceless French VOTs were significantly shorter at post-training than pre-training; (2) for the L1 groups separately, the Mandarin-L1 group produced significantly shorter French VOTs over training in both voiced and voiceless plosives, whereas the English-L1 group did so for voiceless but not for voiced plosives; (3) there were no changes over training for English VOT; and (4) there were substantial individual differences.

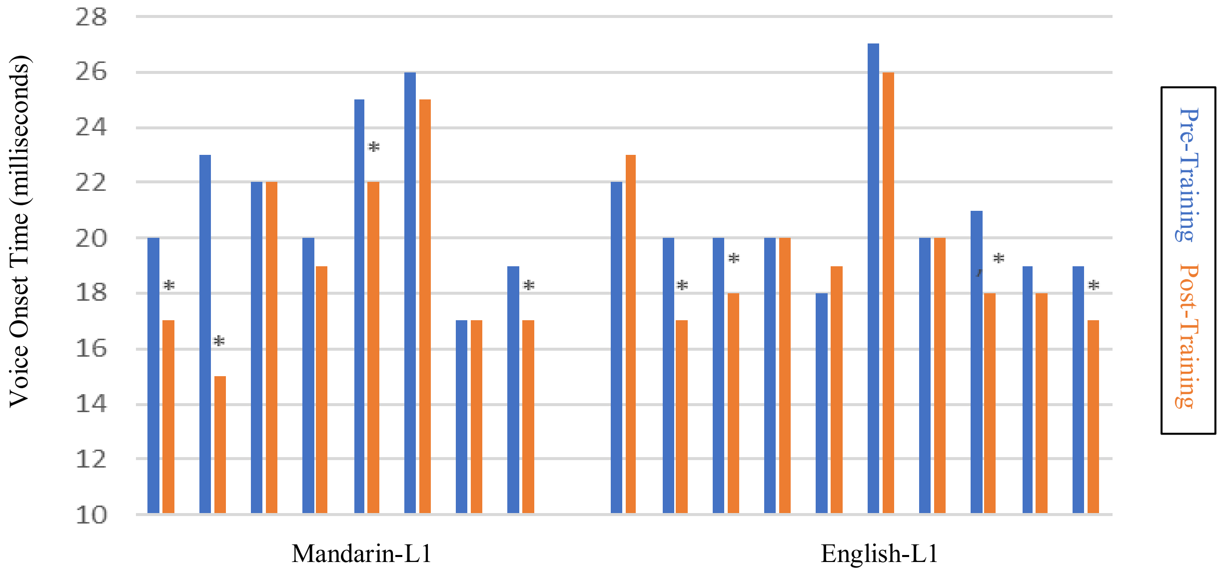


Figure 2. *French Voiced VOT Change per Participant*

The average VOT duration (in msec) for voiced tokens between pre and post training for each participant, separated by L1 group. Significant improvements (in paired sample t-tests) are marked.

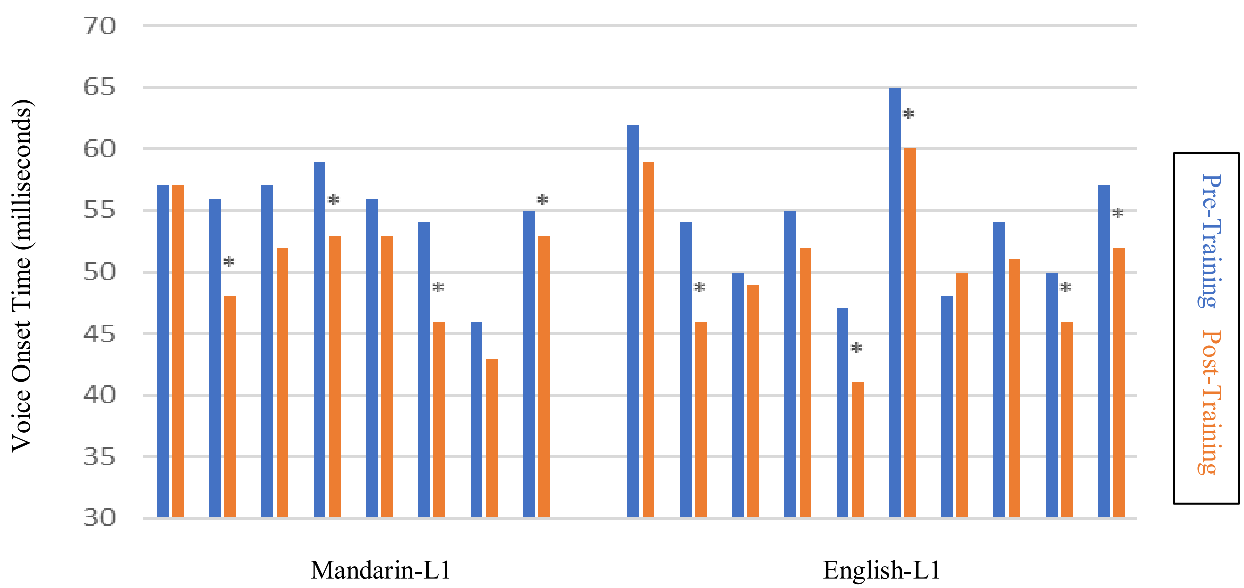


Figure 3. *French Voiceless VOT Change per Participant*

The average VOT duration (in msec) for voiceless tokens between pre and post training for each participant, separated by L1 group. Significant improvements (in paired sample t-tests) are marked.

Vowel Formant Analysis

The French vowel formant frequencies at pre and post training for the 2 L1 groups separately are listed in **Table 3** of the **Supplementary Materials** (section **III**), along with the French reference values according to gender.

To visualise changes in French vowel formants over time, **Figures 4-7** show the position of each vowel at pre and post training in relation to their French and English references on F1 vs F2 graph space, separated by both gender and L1 group. In each of the 4 figures it is possible to observe changes over training, particularly for /i/ production (circled), becoming closer to the French reference value over time. Using these figures, we calculated the Euclidean Distance between each participants' pre and post training average vowel production and corresponding French reference for each vowel separately. From this, we calculated the change in Euclidean Distance over time (pre-training minus post-training), whereby a positive score indicated improvement since a greater Euclidean Distance from the French reference indicates less 'French-like' production (see **Table 5**).

Only /i/ production showed changes over time. We ran 3-way univariate ANOVAs for each formant (F1 and F2) of each vowel (/i/, /u/, /a/, /oe/, /e/, / ε /) with Time (2 levels: pre or post-training), L1 Group (2 levels: Mandarin or English), and Gender (2 levels: male or female) as factors. Results showed significant main effects for Gender for every comparison (all *p*'s <.016) with formant values being higher for women than men, except for F1 for /u/. There were significant main effects of L1 Group for the F2 of /u/ ($F(1, 24) = 21.86$, $p = <.001$, $\eta p^2 = 0.477$), the F2 of /a/ ($F(1, 24) = 4.95$, $p = .036$, $\eta p^2 = 0.171$), and the F2 of /oe/ ($F(1, 24) = 25.24$, $p = <.001$, $\eta p^2 = 0.513$. There were no significant main effects for Time except for the F2 of /i/ ($F(1, 24) = 9.52$, $p = .005$, $\eta p^2 = 0.284$). There were no significant 2-way or 3-way interactions.

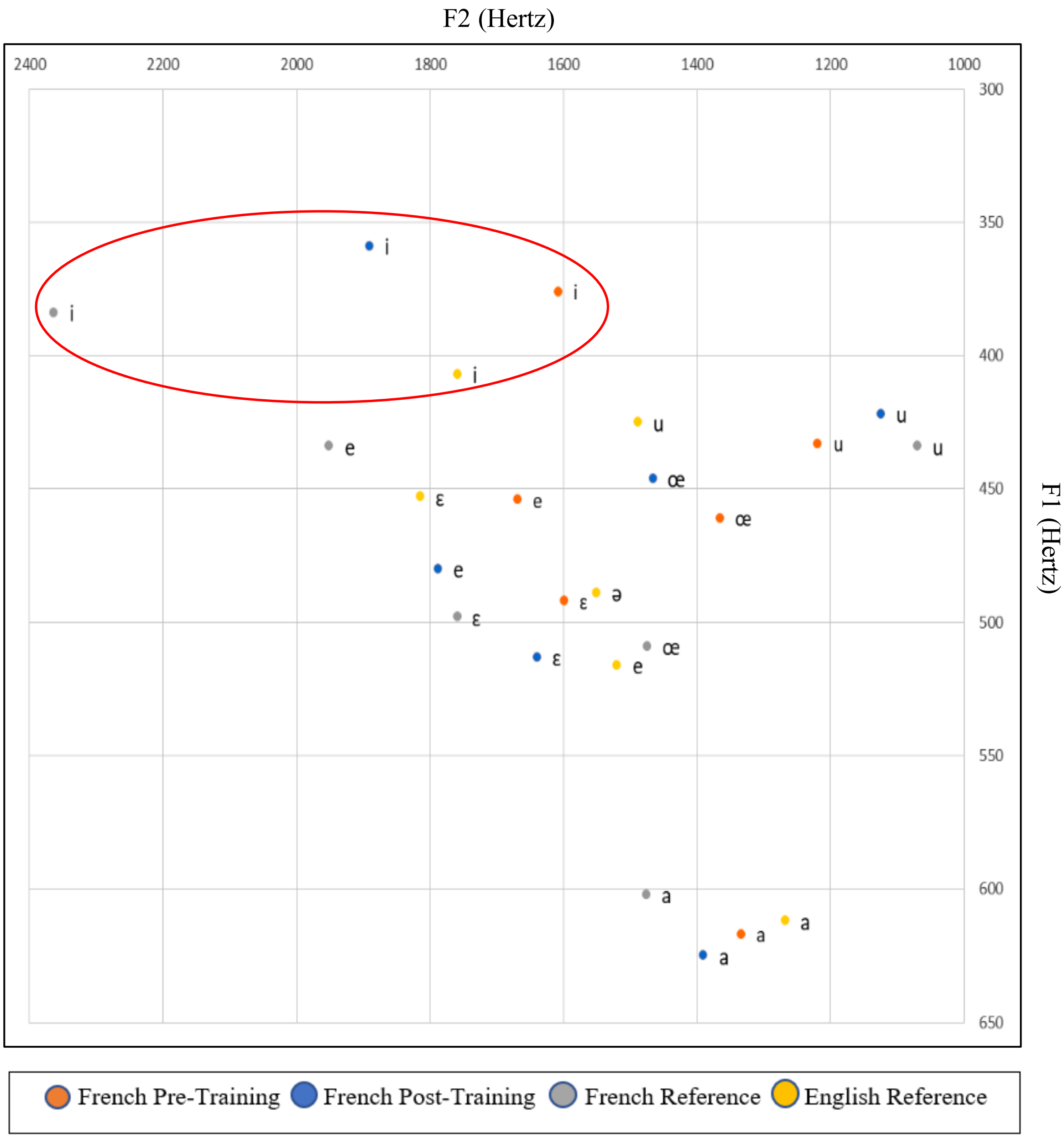


Figure 4. *Mandarin-L1 Male Vowel Production*

The average position of each vowel on F1 vs F2 graph space at pre and post-training for Mandarin-L1 male participants (n=2) in relation to French and English references. It is possible to see improvement over time especially in the /i/ production (circled), which moves closer to the French reference value.

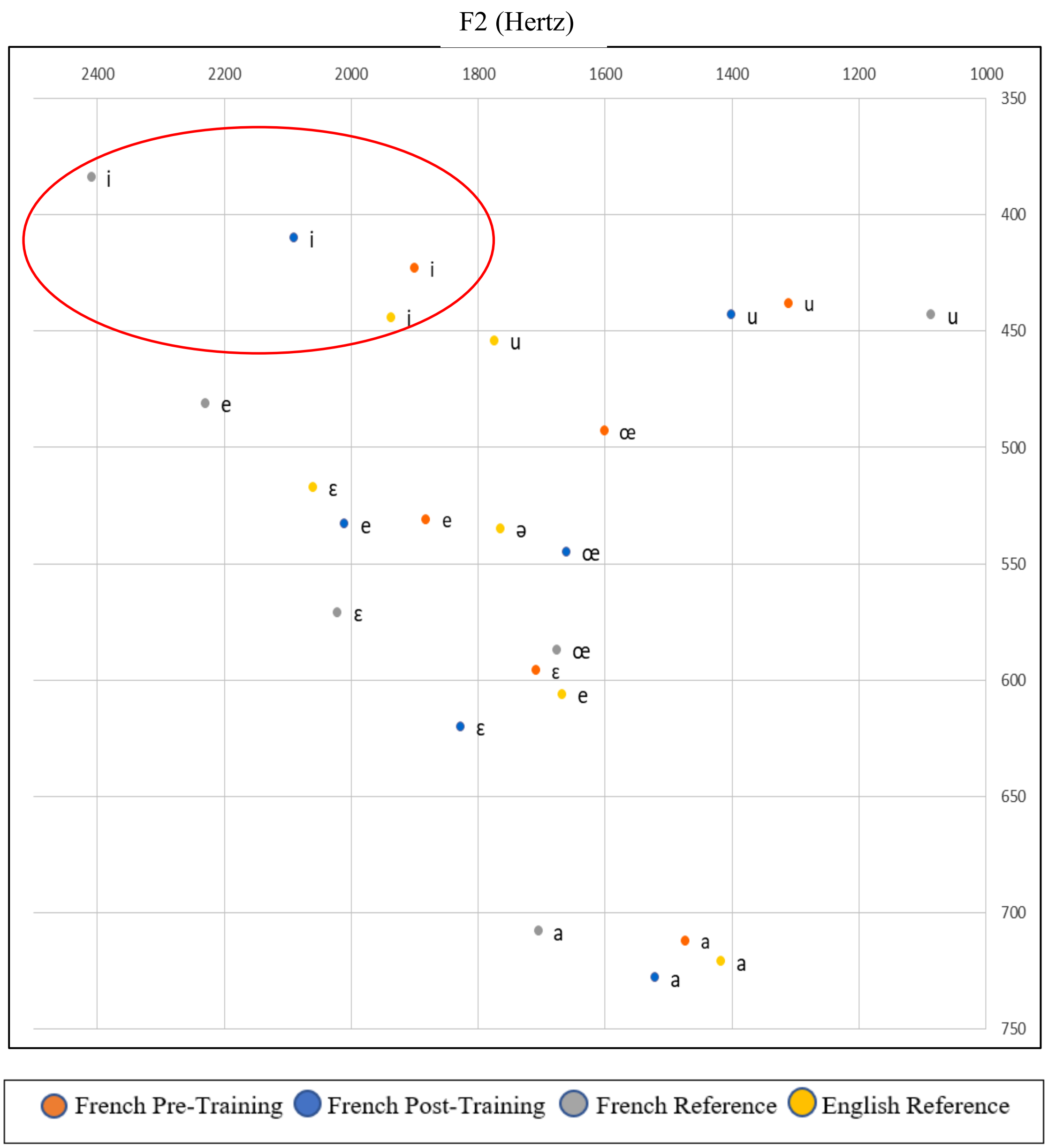


Figure 5. *Mandarin-L1 Female Vowel Production*

The average position of each vowel on F1 vs F2 graph space at pre and post-training for Mandarin-L1 female participants (n=4) in relation to French and English references. It is possible to see improvement over time especially in the /i/ production (circled), which moves closer to the French reference value.

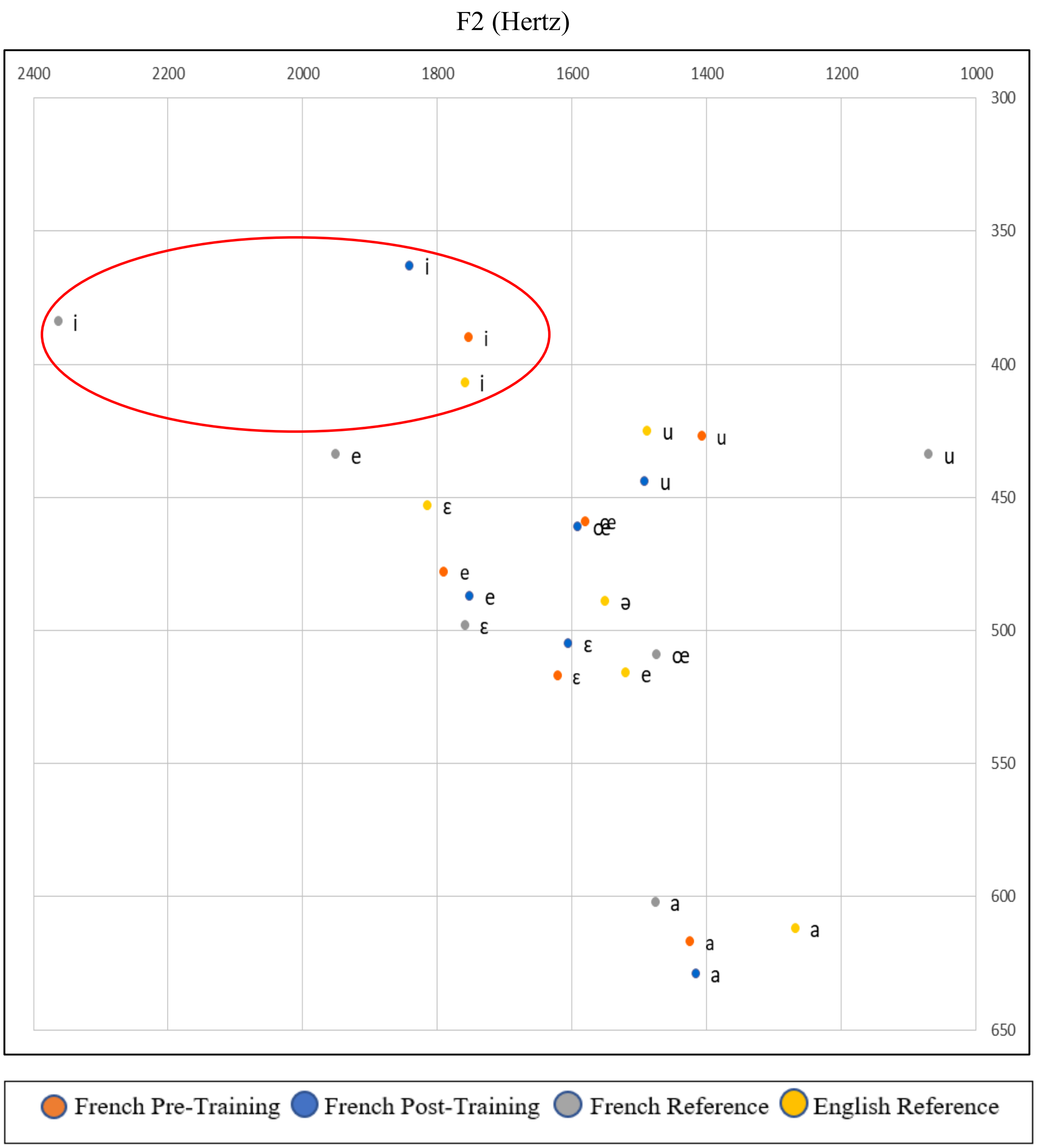


Figure 6. *English-L1 Male Vowel Production*

The average position of each vowel on F1 vs F2 graph space at pre and post-training for English-L1 male participants (n=4) in relation to French and English references. It is possible to see improvement over time especially in the /i/ production (circled), which moves closer to the French reference value.

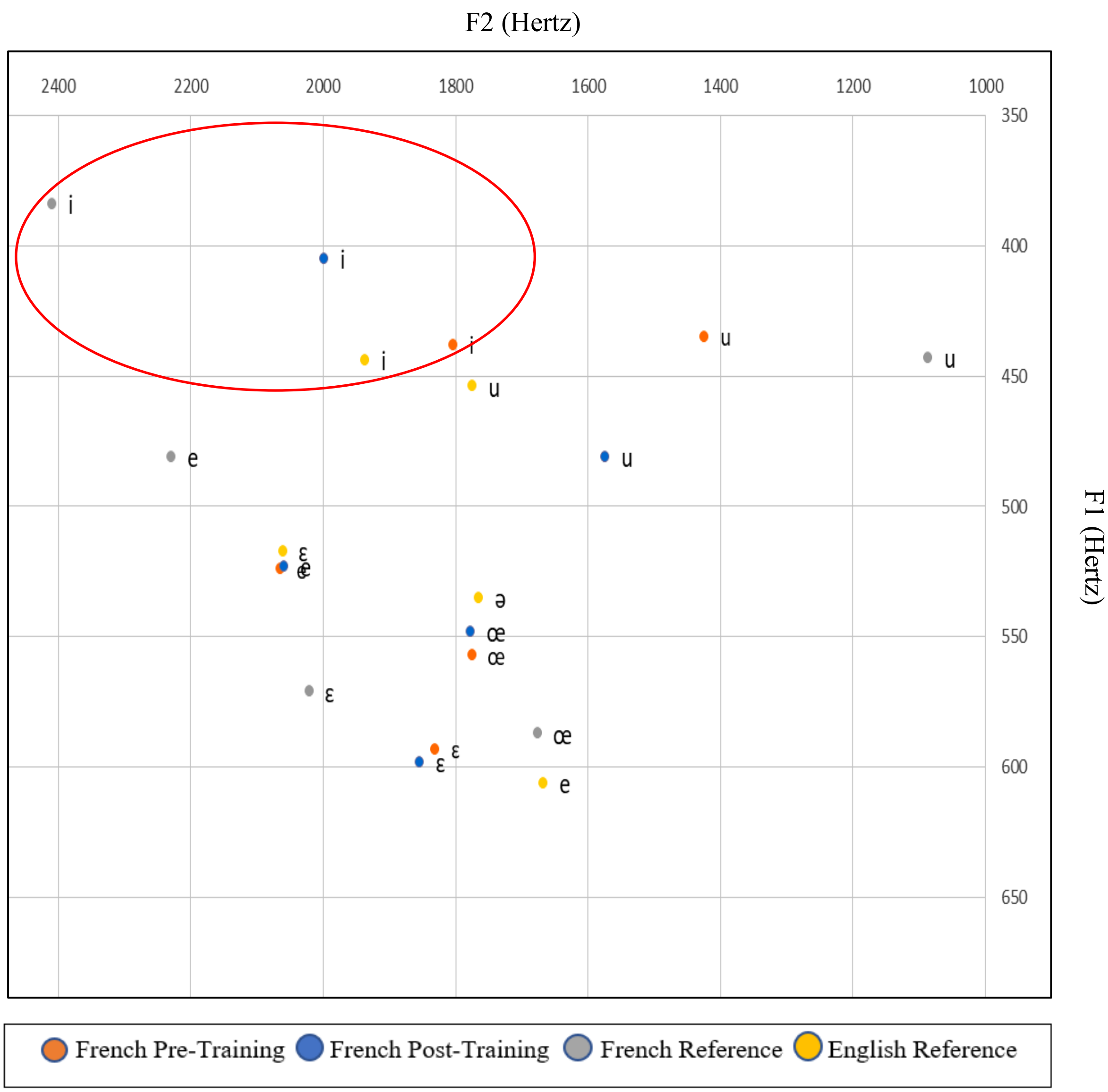


Figure 7. *English-L1 Female Vowel Production*

The average position of each vowel on F1 vs F2 graph space at pre and post-training for English-L1 female participants (n=5) in relation to French and English references. It is possible to see improvement over time especially in the /i/ production (circled), which moves closer to the French reference value.

The Euclidean Distance between /i/ production and the French /i/ reference significantly improved on 'French-ness' over time. Paired sample t-tests indicated that there was significant improvement for all participants combined (t (14) = 3.42, p = .004) and the English-L1 group (t (8) = 2.63, p = .030), but not for the Mandarin-L1 group (p = .103). There was no significant improvement in the 'French-ness' of any of the other vowels for groups combined (all p's > .128) (See **Table 5**). There were some differences between groups in vowel production changes. Specifically, the Mandarin-L1 group alone significantly improved on /oe/ (t (5) = 3.54, p = .017) and /e/ production (t (5) = 3.46, p = .018). The Mandarin-L1 group improved significantly more than the English-L1 group on /e/ production (t (13) = 2.24, p = .043) (See **Table 5**).

Finally, we ran the same ANOVA as for the French analysis comparing English vowel formants (/i/, /u/, /a/, /e/, / ə /, /ɛ/) at pre- and post training. There were no significant effects of Time. The English vowel formant values at pre and post training for English-L1 and Mandarin-L1 participants are listed in **Table 4** in the **Supplementary Materials** (section **III**).

Table 5. French Vowel Changes

	/i/	/u/	/a/	/oe/	/e/	/ ɛ /	mean
English-L1	145* *(156)*	-93 *(135)*	-46 *(86)*	-4 *(84)*	-18 *(136)*	9 *(66)*	-1 *(48)*
Mandarin-L1	139 *(156)*	-72 *(165)*	46 *(110)*	60* *(38)*	120* *(77)*	58 *(53)*	58 *(63)*
Combined	143* *(162)*	-84 *(154)*	-9 *(110)*	22 *(79)*	37 *(139)*	28 *(68)*	23 *(64)*
t Statistic	*0.07*	*0.27*	*1.74*	*1.99*	*2.24*	*1.52*	*2.06*
P Value	*.943*	*.791*	*.106*	*.071*	*.043**	*.153*	*0.060*

Table 5: The change in Euclidean Distance between average vowel production and the corresponding French reference at pre and post training, to give a measure of improvement in vowel production. Positive values indicate improvement. The t statistic and p value refer to the group comparisons by independent samples t-tests. Mean *(SD)*.

In summary, we found that: (1) for the L1 groups combined, only French /i/ production significantly improved; (2) for the L1 groups separately, the English-L1 group but not the Mandarin-L1 group improved on French /i/ production, whereas the Mandarin-L1 group but not the English-L1 group improved on French /oe/ and /e/ production; (3) there were no training effects in any English vowel for either the L1 groups combined or separate; and (4) there are substantial individual differences across all these measures.

5.1.2 Speech and Language Tasks

1) Words per Minute (reading)

In French, both L1 groups improved significantly between pre and post-training. English-L1 participants performed significantly better than Mandarin-L1 participants at post-training, but not at pre-training. The extent of improvement was not significantly different between the groups (see **Table 6**). In English, there were no learning effects for either group (see **Table 7**).

Table 6. French Words per Minute

	Pre-Training	Post-Training	Change	*t Statistic*	*P Value*
English-L1	80 *(19)*	100 *(11)*	20 *(16)*	*-3.83*	*.004**
Mandarin-L1	66 *(17)*	80 *(14)*	14 *(5)*	*-7.35*	*<.001**
t Statistic	*1.63*	*3.4*	*1.02*	_	_
P Value	*.123*	*.004**	*.325*	_	_

Mean *(SD)*

Table 7. English Words per Minute

	Pre-Training	Post-Training	Change	*t Statistic*	*P Value*
English-L1	197 *(16)*	183 *(20)*	-14 *(22)*	*-2.04*	*.072*
Mandarin-L1	134 *(28)*	132 *(17)*	-2 *(20)*	*-0.30*	*.771*

Mean *(SD)*

2) Picture Description Task

The 7 measurements used to quantify proficiency in this task were described in **Table 3**. In French, paired samples t-tests showed that the English-L1 group improved significantly in 4 measures - Vocabulary Index (t (9) = -4.81, p = .001), Total Words (t (9) = -3.25, p = .010), Grammar Index (t (9) = -2.71, p = .024), and Articulation Rate (t (9) = -2.64, p = .027). The Mandarin-L1 group also improved significantly in 4 measures - Filled Pauses (t (7) = 2.83, p = .025), Vocabulary Index (t (7) = -3.33, p = .013), Total Words (t (7) = -2.87, p = .024), and Articulation Rate (t (7) = -3.22, p = .015). There were no significant L1 group differences in performance or improvement for any measures (see **Table 8**). The results of the accent ratings (averaged across both independent raters) showed no significant improvement for either L1 group (p's > 0.74).

The results of the English Picture Description Task are shown in **Table 9**. For Mandarin-L1 participants, there was significant improvement in Total Words (t (7) = -3.62, p = .008), Vocabulary Index ((t (7) = -5.03, p = .002), and Grammar Index (t (7) = -3.31, p = .013). For English-L1 participants, there was no significant improvement in any measure.

In summary, for the Picture Description Task, we found that: (1) the Mandarin-L1 group improved significantly in French on 4 measures between pre and post-training: Total Words, Vocabulary Index, Articulation Rate, and Filled Pauses; (2) the English-L1 group also improved significantly in French on 4 measures: Total Words, Vocabulary Index, Grammar Index, and Articulation Rate; (3) the L1 groups did not differ significantly in their performance or extent of French improvement on any measures; and (4) for the Mandarin-L1 group only there were training effects in English on 3 measures: Total Words, Vocabulary Index, and Grammar Index.

Table 8. French Picture Description

	Pre-Training		Post-Training		Change	
	English-L1	Mandarin-L1	English-L1	Mandarin-L1	English-L1	Mandarin-L1
Total Words	68 *(27)*	56 *(16)*	94 *(24)*	80 *(25)*	26* *(25)*	24* *(24)*
Vocabulary Index	33 *(13)*	28 *(10)*	49 *(11)*	41 *(11)*	16* *(10)*	13* *(11)*
Grammar Index	13 *(19)*	11 *(12)*	33 *(20)*	27 *(22)*	20* *(23)*	16 *(22)*
Articulation Rate	0.84 *(0.25)*	0.66 *(0.07)*	0.98 *(0.24)*	0.85 *(0.19)*	0.14* *(0.38)*	0.19* *(0.17)*
Mean Syllables per Run	4.2 *(1.3)*	3.6 *(1.0)*	4.5 *(1.3)*	3.9 *(0.7)*	0.3 *(1.0)*	0.4 *(0.7)*
Silent Pauses	52.9 *(21.1)*	65.9 *(14.3)*	59.1 *(17.2)*	59.6 *(11.2)*	-6.1 *(21.6)*	6.3 *(15.9)*
Filled Pauses	2.61 *(1.79)*	3.8 *(2.35)*	2.23 *(1.55)*	1.46 *(0.55)*	0.38 *(2.39)*	2.39* *(2.38)*
Accent Ratings	2.05 *(0.8)*	2.75 *(0.9)*	2.55 *(1.3)*	2.81 *(1.3)*	0.50 *(0.8)*	0.06 *(1.0)*

Table 8: The scores across time in the measures extracted from the free speech samples. Significant improvements within L1 groups are marked. For Silent Pauses and Filled Pauses, a lower value means better performance. The units are described in **Table 3.** Mean *(SD)*.

Table 9. English Picture Description Task

	Pre-Training		Post-Training		Change	
	English-L1	Mandarin-L1	English-L1	Mandarin-L1	English-L1	Mandarin-L1
Total Words	239 *(78)*	148 *(49)*	263 *(61)*	204 *(31)*	24 *(40)*	56* *(44)*
Vocabulary Index	131 *(43)*	77 *(26)*	135 *(23)*	108 *(19)*	4 *(28)*	32* *(18)*
Grammar Index	48 *(18)*	41 *(16)*	41 *(15)*	55 *(14)*	-7 *(18)*	13* *(11)*
Articulation Rate	3.08 *(0.66)*	2.31 *(0.28)*	3.06 *(0.58)*	2.26 *(0.41)*	-0.02 *(0.41)*	-0.05 *(0.41)*
Mean Syllables per Run	12.4 *(4.4)*	6.1 *(1.1)*	12.5 *(2.7)*	6.4 *(1.1)*	0.1 *(2.9)*	0.4 *(1.2)*
Silent Pauses	28.1 *(9.5)*	31.5 *(8.9)*	41.7 *(13.6)*	42.9 *(11.2)*	-13.6 *(14.2)*	-11.4 *(14.4)*
Filled Pauses	3.3 *(2)*	5.8 *(8.5)*	2.6 *(1.55)*	2.8 *(2.3)*	0.7 *(1.9)*	2.9 *(9)*

Table 9: The scores across time in the measures extracted from the free speech samples. Significant improvements within L1 groups are marked. For Silent Pauses and Filled Pauses, a lower value means better performance. The units are described in **Table 3.** Mean *(SD)*.

3) Sentence Repetition Task

In French, paired samples t-tests showed significant improvement in the average percentage of words accurately repeated (across 24 sentences) for both English-L1 and Mandarin-L1 groups. Between groups, neither the differences in performance or the difference in extent of improvement was significant (see **Table 10**). In English, there were no practice effects for either group and no differences between groups in performance or improvement (see **Table 11**).

Table 10. French Sentence Repetition

	Pre-Training	Post-Training	Change	*t Statistic*	*P Value*
English-L1	28 *(7.6)*	40 *(9.9)*	11.8 *(8.3)*	*-4.55*	*.001**
Mandarin-L1	24 *(7.9)*	32 *(12)*	7.5 *(3.8)*	*-3.55*	*.009**
t Statistic	*1.09*	*1.55*	*1.35*	_	_
P Value	*.292*	*.140*	*.196*	_	_

Mean *(SD)*

Table 11. English Sentence Repetition

	Pre-Training	Post-Training	Change	*t Statistic*	*P Value*
English-L1	99 *(7.3)*	98 *(5.6)*	-1.1 *(6.4)*	*-0.34*	*.735*
Mandarin-L1	95 *(11.1)*	96 *(9.3)*	1.1 *(8.6)*	*0.20*	*.848*
t Statistic	*0.92*	*0.57*	*-0.62*	_	_
P Value	*.371*	*.579*	*.542*	_	_

Mean *(SD)*

4) Verbal Fluency Task

In French, there were no significant differences between L1 groups at either time point and both groups improved significantly (See **Table 12**). In English, there were no significant differences between groups at either time point and the Mandarin-L1 group but not the English-L1 group improved significantly (see **Table 13**).

Table 12. French Letter Fluency

	Pre-Training	Post-Training	*t Statistic*	*P Value*
English-L1	19.9 *(6.5)*	22.8 *(6.5)*	*2.42*	*.038**
Mandarin-L1	15.6 *(4.8)*	19.1 *(5.3)*	*2.79*	*.027**
t Statistic	*1.56*	*1.30*	_	_
P Value	*.139*	*.212*	_	_

Mean *(SD)*

Table 13. English Letter Fluency

	Pre-Training	Post-Training	*t Statistic*	*P Value*
English-L1	43.4 *(9.3)*	45.3 *(10.8)*	*0.68*	*.511*
Mandarin-L1	35.5 *(6.8)*	38.5 *(6.4)*	*2.85*	*.025**
t Statistic	*2.01*	*1.57*	_	_
P Value	*.062*	*.136*	_	_

Mean *(SD)*

5) Hindi Discrimination Task

There were no significant differences between L1 groups at either time point and no significant improvement over time for either group (see **Table 14**).

Table 14. Hindi Discrimination

	Pre-Training	Post-Training	*t Statistic*	*P Value*
English-L1	326.8 *(40.7)*	344 *(72.9)*	*0.99*	*.350*
Mandarin-L1	315 *(29.9)*	308.3 *(22.4)*	*-0.62*	*.554*
t Statistic	*0.72*	*1.34*	–	–
P Value	*.480*	*.199*	–	–

Mean *(SD)*

5.2 Voxel Based Morphometry Results

5.2.1 Neuroplasticity

We compared GMV between pre-training and post-training for both groups separately, and for the groups combined. We found no significant effects of GMV neuroplasticity over time either in the whole-brain analysis or in the voxel-specified analyses (all p's > .085). The results of the voxel-specified analyses are shown in the **Supplementary Materials** (Section **IV**).

5.2.2 Biomarkers

The measures which showed significant improvement across our L1 groups combined and which showed no change in the equivalent English task were: (1) VOT (we used the 'Combined' change, whereby a lower score indicated more improvement); (2) /i/ Vowel Production; (3) Articulation Rate; (4) Sentence Repetition (accuracy); and (5) Words per Minute (reading).

1) <u>VOT</u>

VOT change was significantly negatively correlated with pre-training GMV in the right cerebellum, lobule VIIIb. The data are illustrated in **Figures 8a** and **8b**.

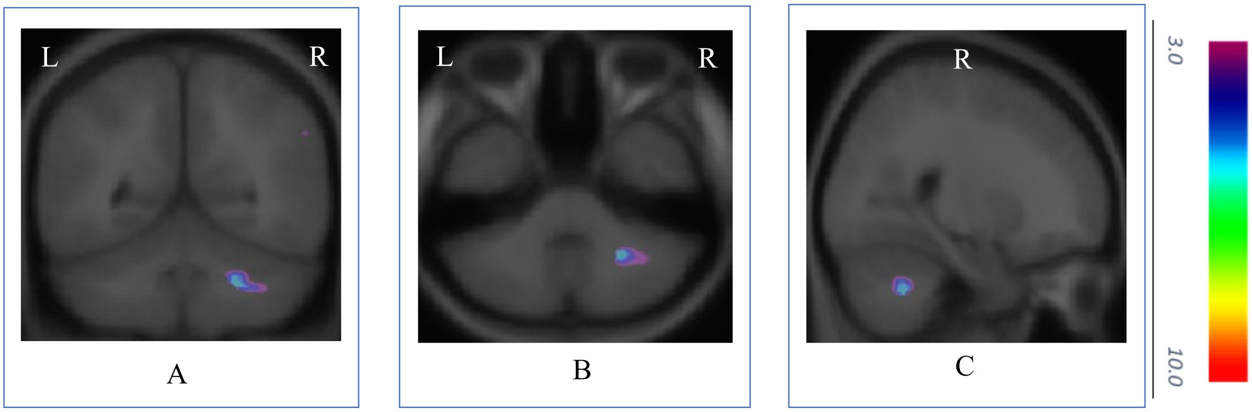


Figure 8a. *Imaging results for VOT change and the right cerebellum*: A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): 24, -54, -39. *t* value: 5.68.

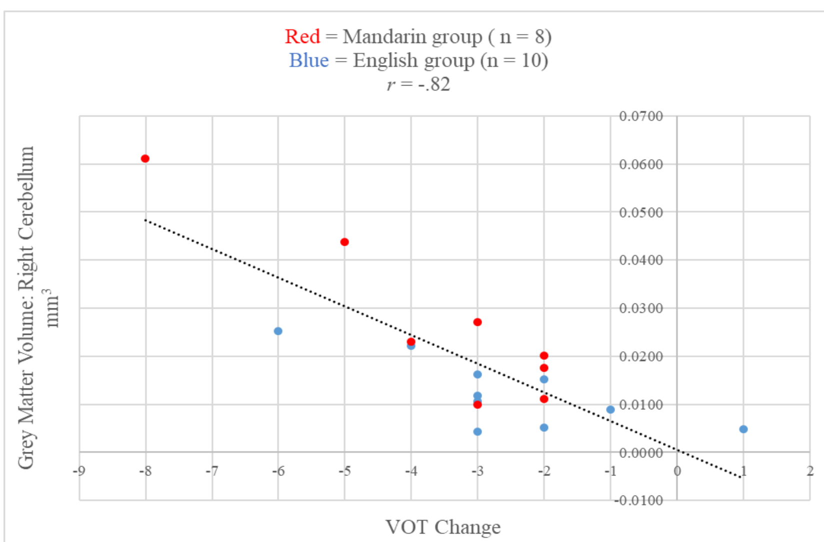


Figure 8b. *Correlation between pre-training right cerebellum GMV and VOT change.* (*r* (16) = -.82, *p* = <.001). A lower VOT change score indicates more improvement.

VOT change was also significantly positively correlated with pre-training GMV in 3 brain areas: the left dorsal and the left ventral premotor cortex (PMC) (BA 6), and the left dorsomedial prefrontal cortex (dmPFC) (BA 9), illustrated respectively in **Figures 9a**,**b** to **11a**,**b**.

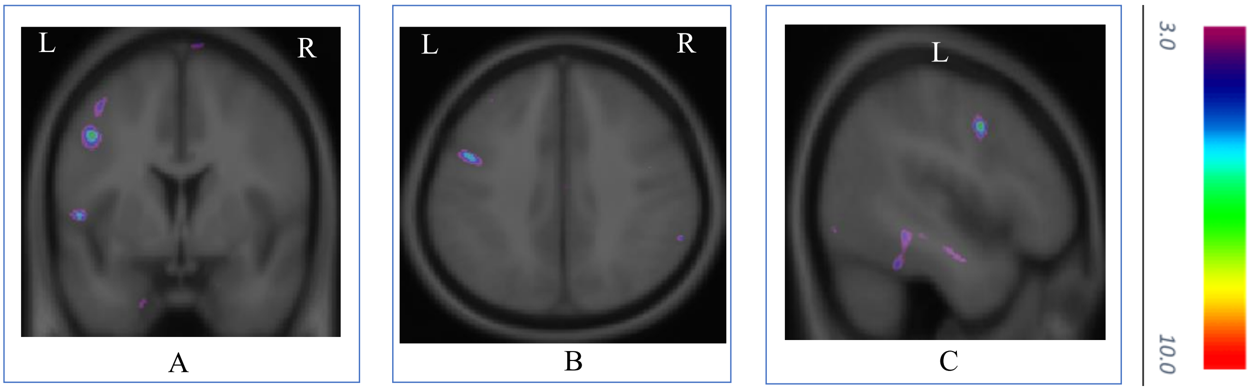


Figure 9a. *Imaging results for VOT change and the left dorsal PMC*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -45, 2, 39. *t* value: 6.59.

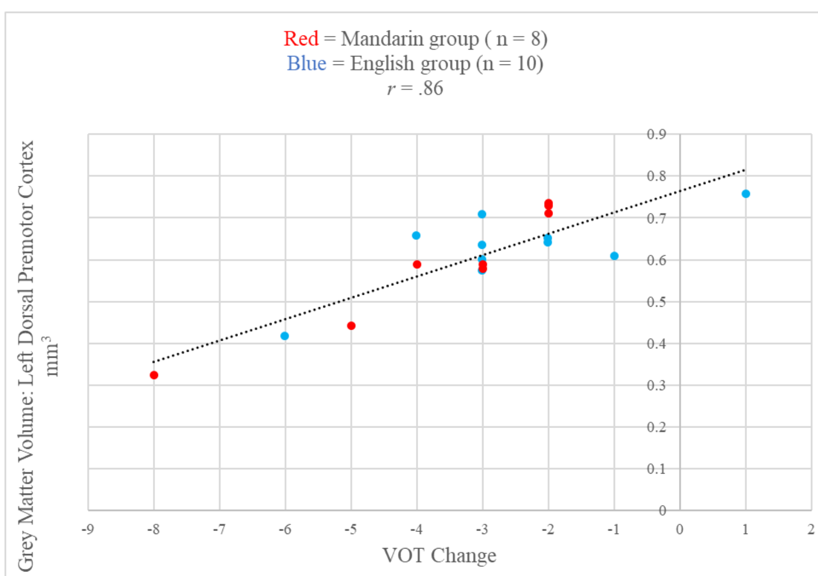


Figure 9b. *Correlation between pre-training left dorsal premotor cortex GMV and VOT change.* (*r* (16) = .86, *p* = <.001). A lower VOT change score indicates more improvement.

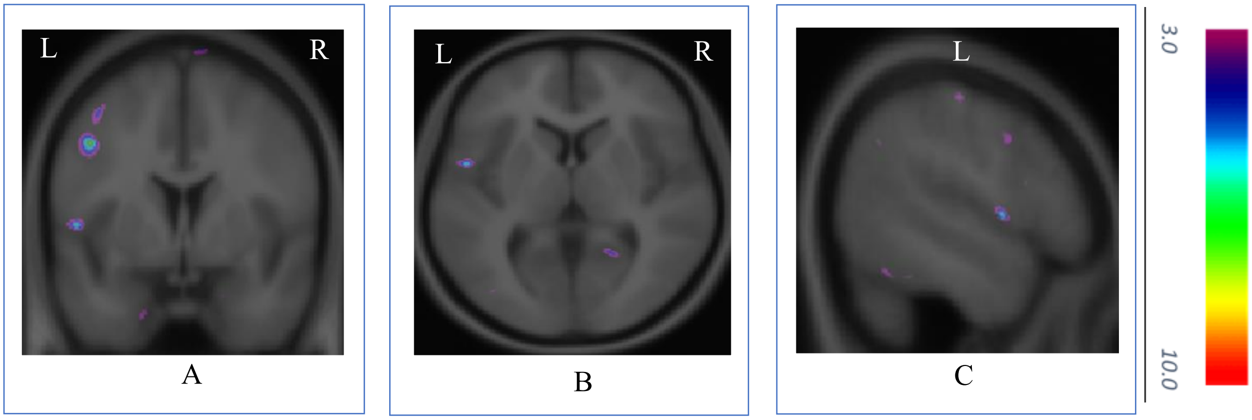


Figure 10a. *Imaging results for VOT change and the left ventral PMC*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -52, 2, 4. *t* value: 5.70.

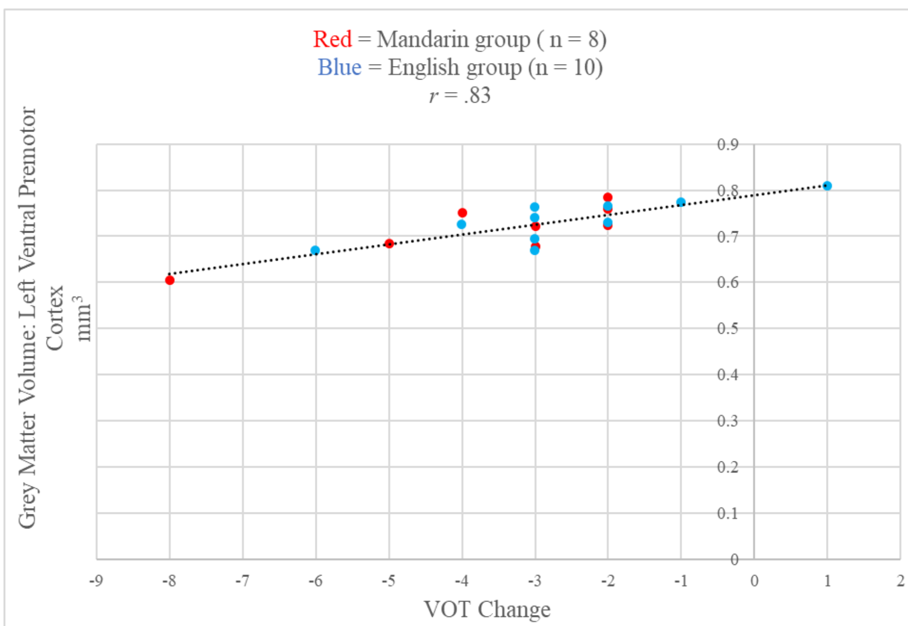


Figure 10b. *Correlation between pre-training left ventral premotor cortex GMV and VOT change.* (*r* (16) = .83, *p* = <.001). A lower VOT change score indicates more improvement.

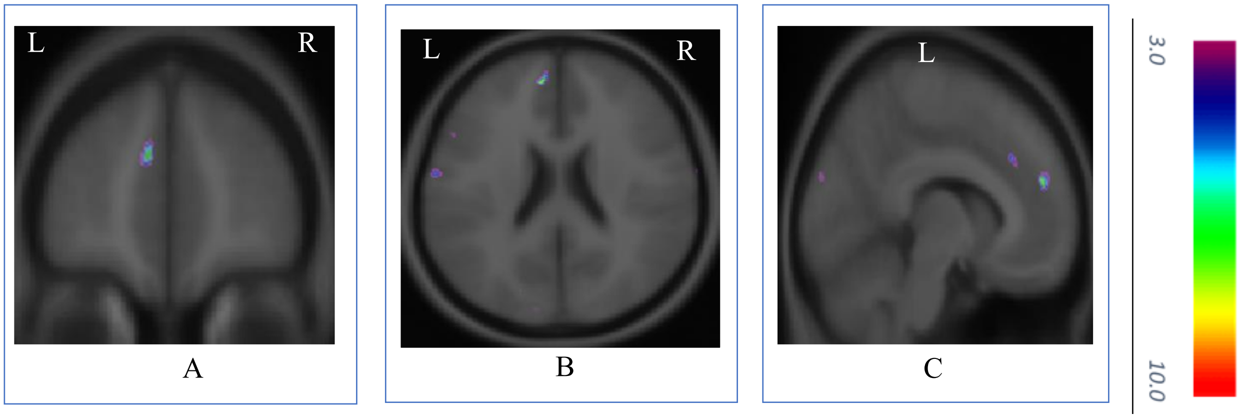


Figure 11a. *Imaging results for VOT change and the left dorsomedial PFC.* A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -9, 49, 23. *t* value: 7.77.

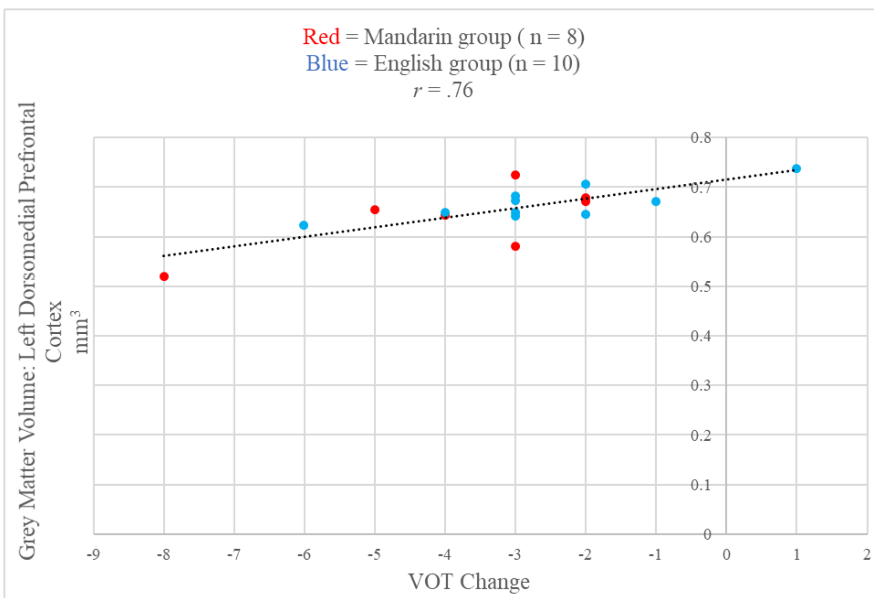


Figure 11b. *Correlation between pre-training left dorsomedial PFC GMV and VOT change. r* (16) = .76, *p* = <.001). A lower VOT change score indicates more improvement.

2) <u>/i/ Vowel Change</u>

/i/ vowel change was significantly correlated with pre-training GMV in the left and the right cerebellum, lobule 8b in both hemispheres, illustrated respectively in **Figures 12 a,b** to **13 a,b**.

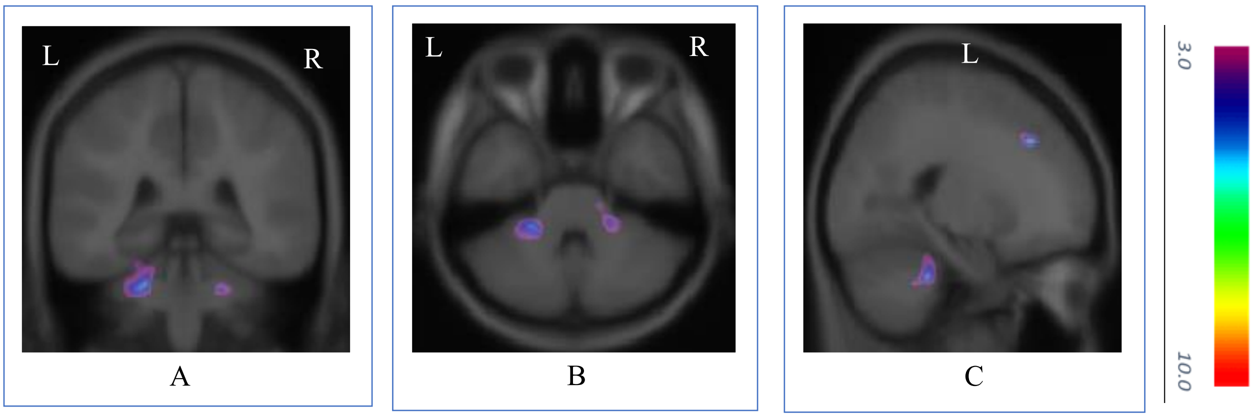


Figure 12a. *Imaging results for /i/ vowel change and the left cerebellum.* A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -20, -39, -32. *t* value: 4.49.

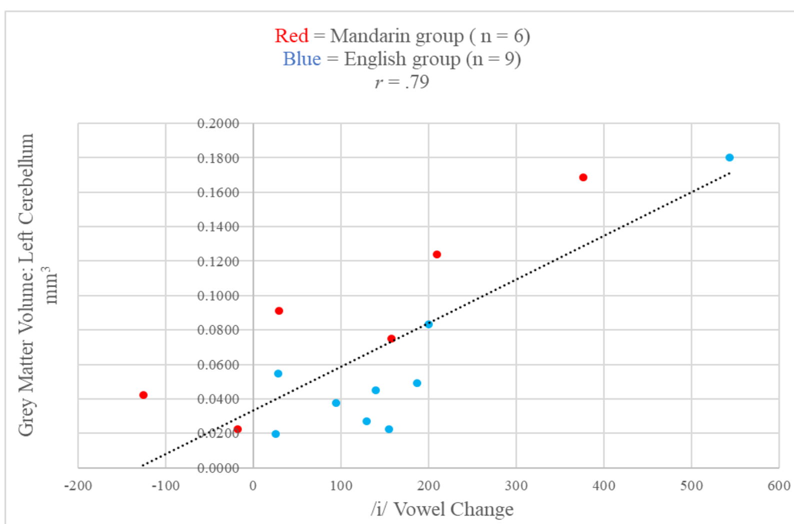


Figure 12b. *Correlation between pre-training left cerebellum GMV and /i/ change*. (*r* (13) = .79, *p* = <.001).

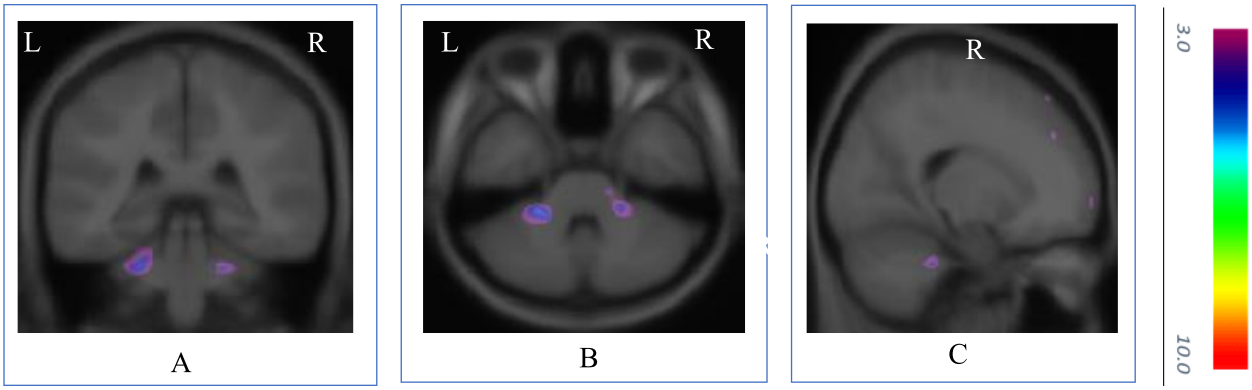


Figure 13a. *Imaging results for /i/ vowel change and the right cerebellum*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): 19, -36, -34. *t* value: 4.15.

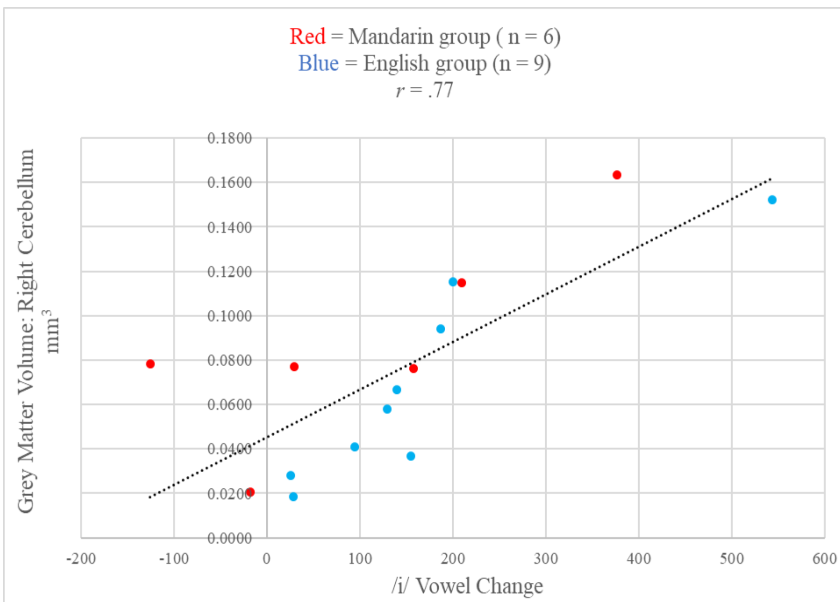


Figure 13b. *Correlation between pre-training right cerebellum GMV and /i/ change.* (r (13) = .77, p = <.001).

3) Articulation Rate

Articulation Rate change was significantly positively correlated with pre-training GMV in 2 brain regions: the left caudate and the supramarginal gyrus (SMG) of the right inferior parietal lobule (IPL) (BA 40), illustrated respectively in **Figures 14 a,b** to **15 a,b**.

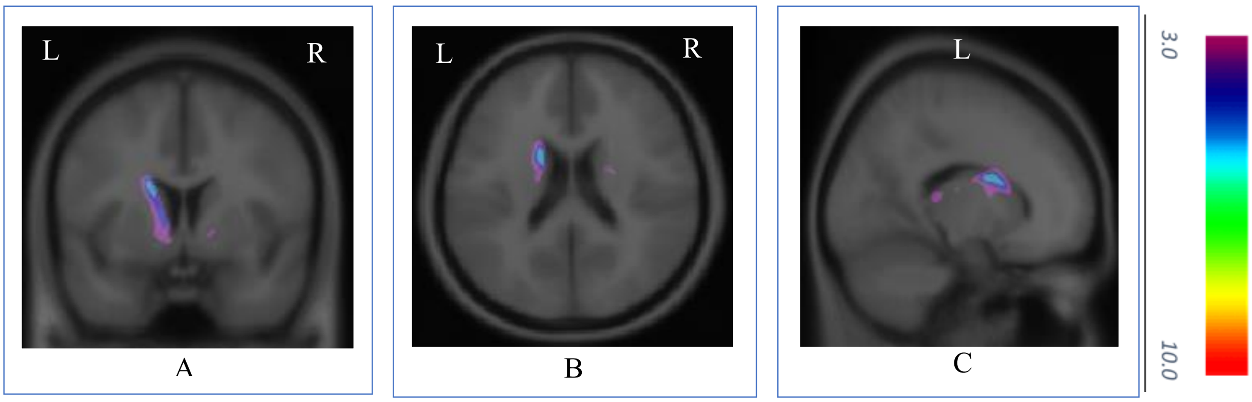


Figure 14a. *Imaging results for articulation rate change and the left caudate*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -16, 4, 20. *t* value: 5.71.

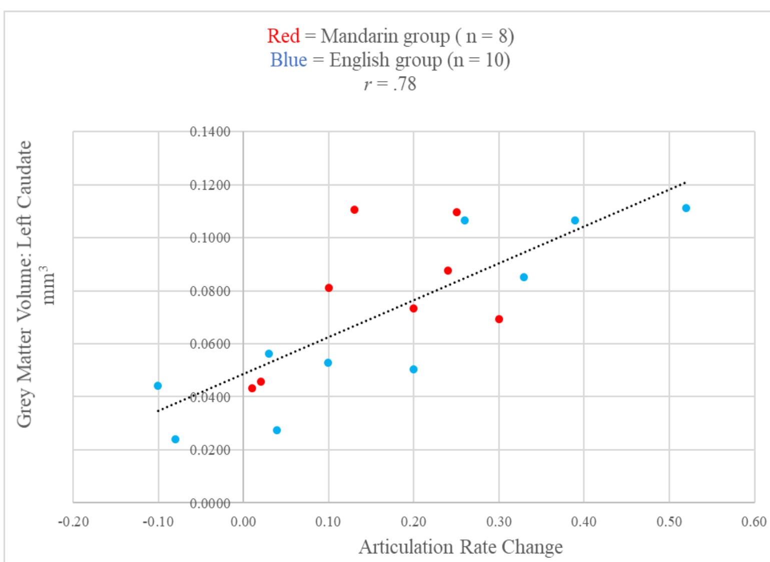


Figure 14b. *Correlation between pre-training left caudate GMV and articulation rate change.* (r (16) = .78, p = <.001).

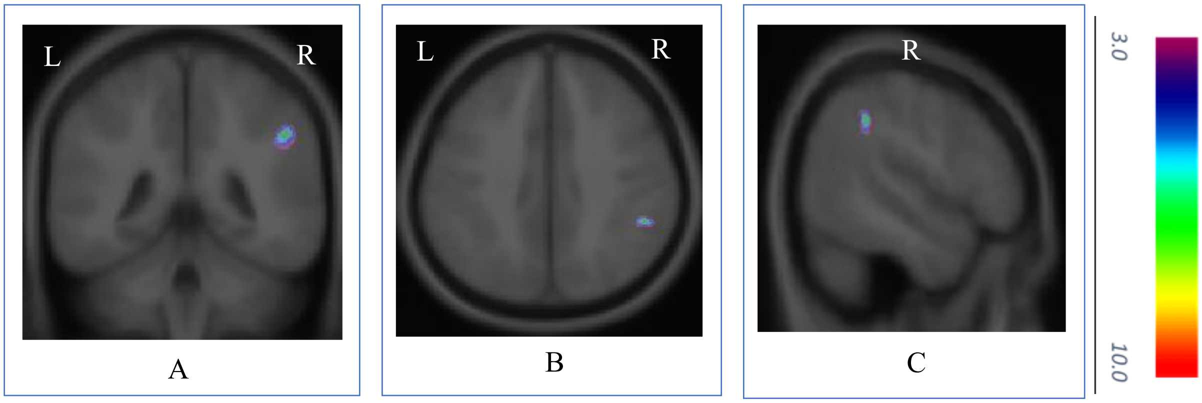


Figure 15a. *Imaging results for articulation rate change and the right IPL*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): 50, -43, 38. *t* value: 6.46.

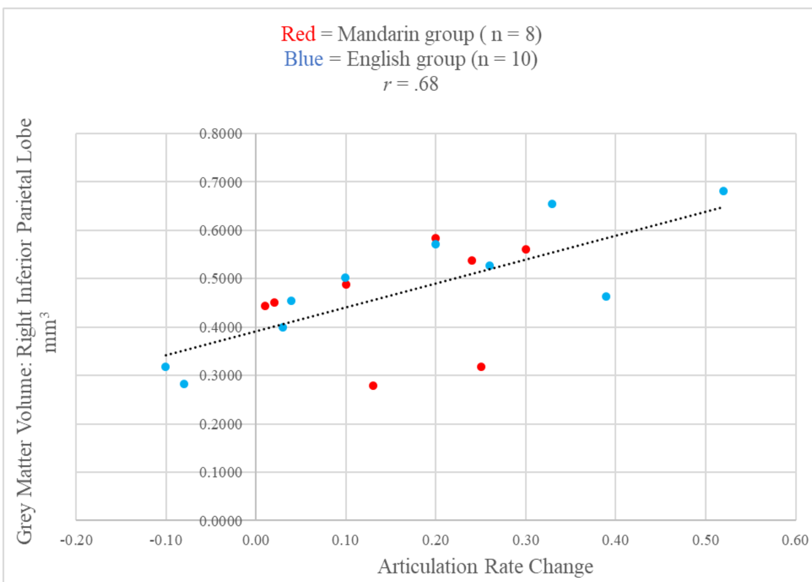


Figure 15b. *Correlation between pre-training right IPL GMV and articulation rate change*. (*r* (16) = .68, *p* = .002).

Additionally, there was a significant positive correlation between Articulation Rate change and pre-training GMV in the left dorsal anterior cingulate (ACC) (BA 32) for the Mandarin-L1 participants only (See **Figures 16a,b**).

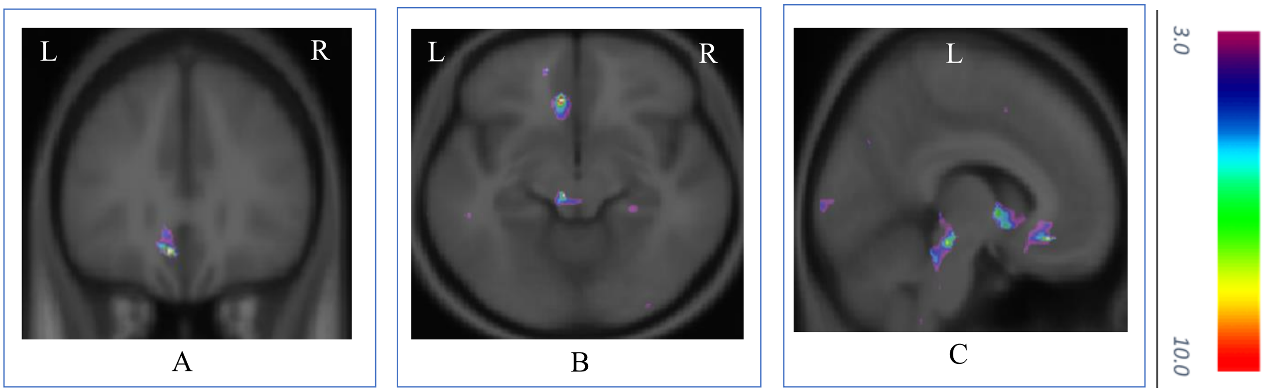


Figure 16a. *Imaging results for articulation rate change and the left dorsal ACC in the Mandarin-L1 group.* A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -7, 31, -11. *t* value: 10.0.

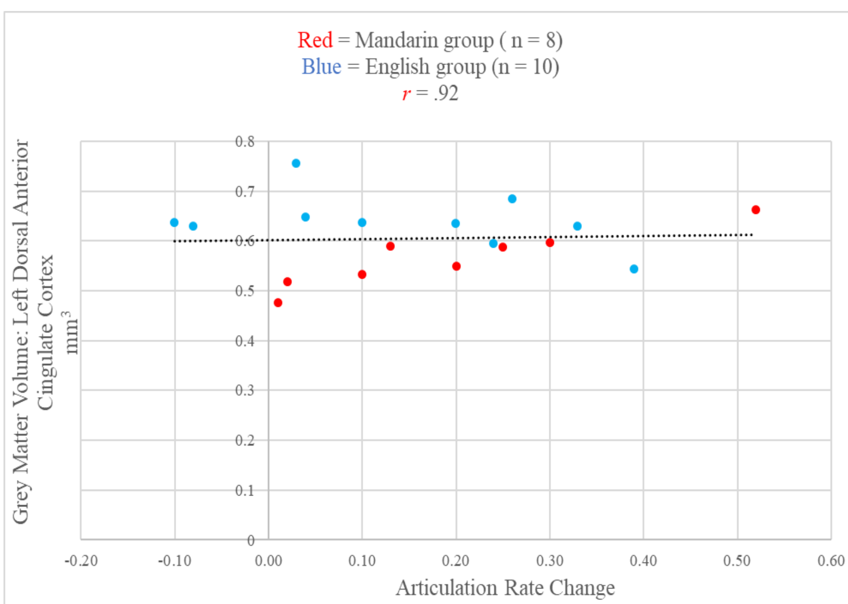


Figure 16b. *Correlation between pre-training left dorsal ACC GMV and articulation rate change for all participants.* (r (6) = .92, p = .001) for the Mandarin-L1 (red) group.

4) Sentence Repetition (accuracy)

Sentence Repetition accuracy change was significantly positively correlated with pre-training GMV in 2 brain regions: the left globus pallidus (GP) and the left middle temporal gyrus (MTG) (BA 21), illustrated in **Figures 17 a,b** to **18 a,b** respectively.

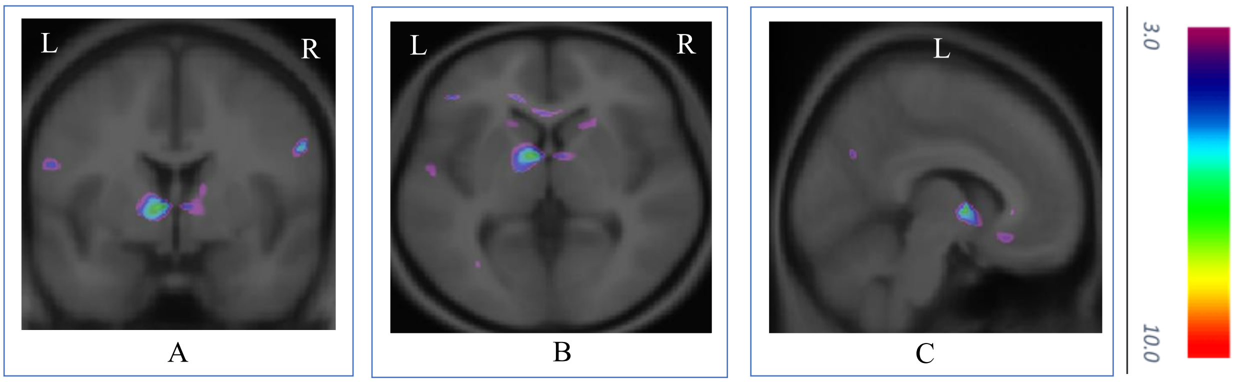


Figure 17a. *Imaging results for sentence repetition change and the left GP*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -7, 0, 2. *t* value: 6.71

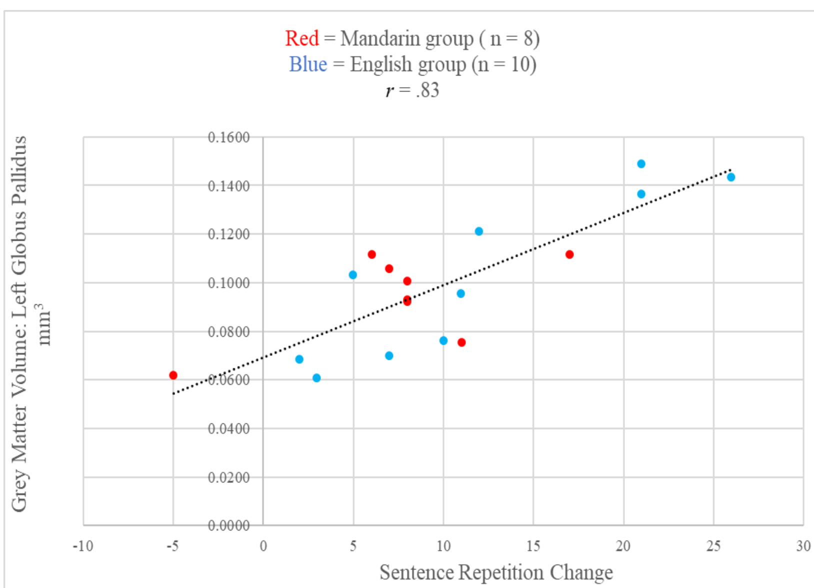


Figure 17b. *Correlation between pre-training left globus pallidus GMV and sentence repetition change.* (*r* (16) = .83, *p* = <.001).

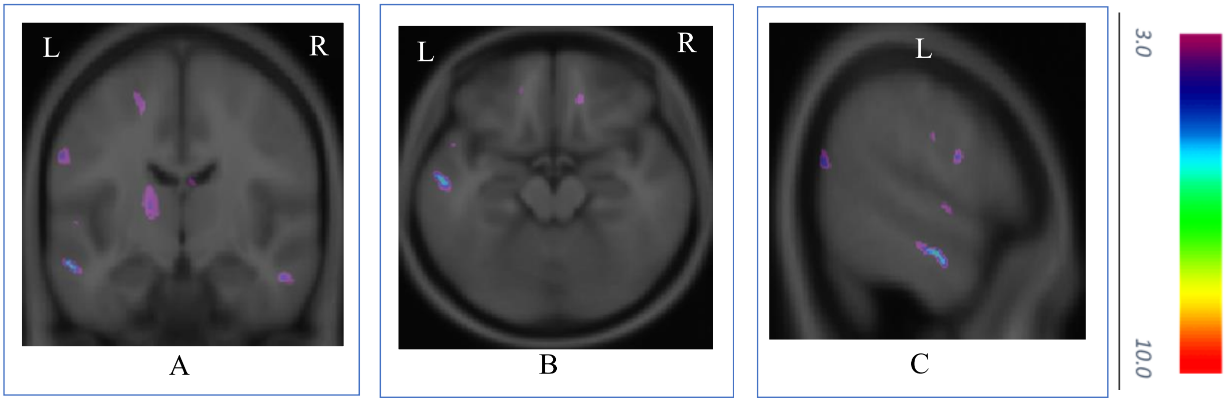


Figure 18a. *Imaging results for sentence repetition change and the left MTG*. A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -56, -12, -17. *t* value: 5.74

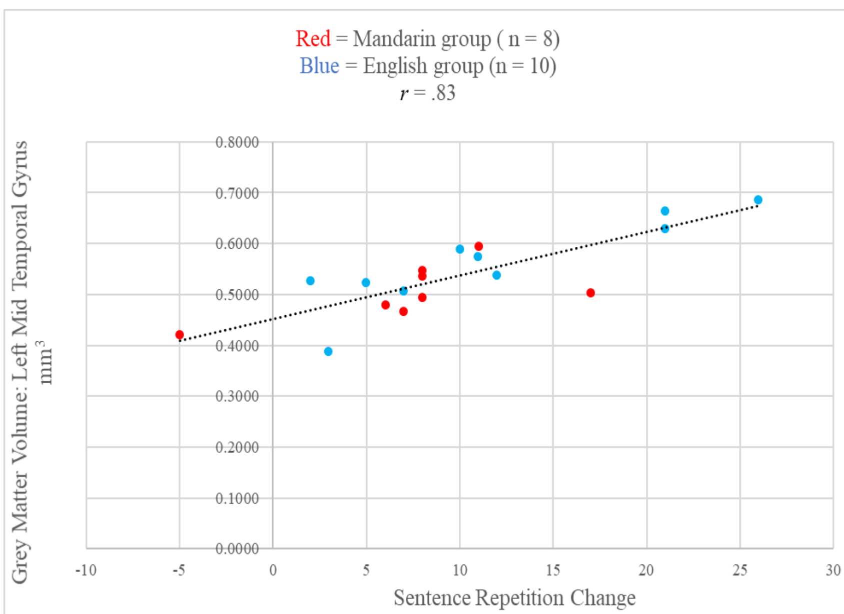


Figure 18b. *Correlation between pre-training left MTG GMV and sentence repetition change*. (*r* (16) = .83, *p* = <.001).

5) <u>Words per Minute (reading)</u>

Words per Minute change was significantly correlated with pre-training GMV in the left middle occipital gyrus (MOG) (BA 18/19), illustrated in **Figures 19 a,b**.

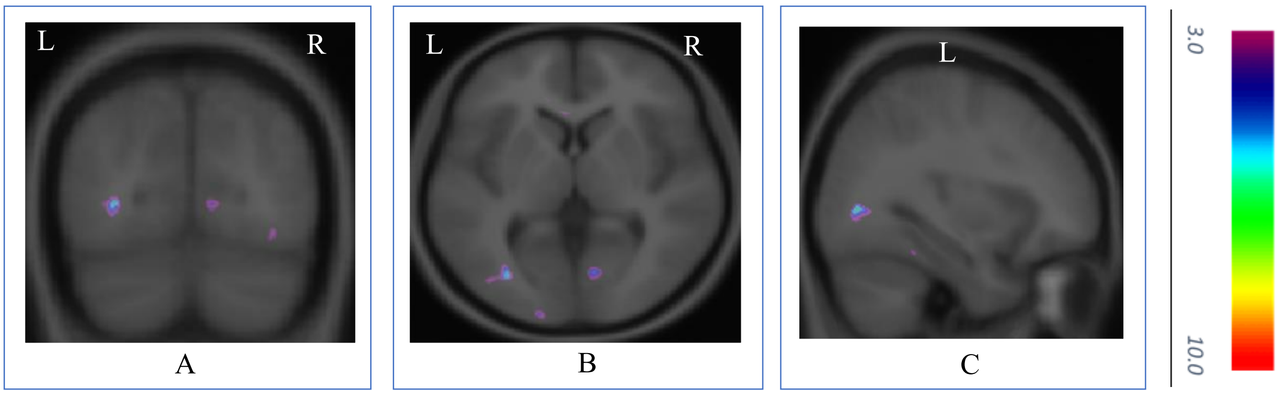


Figure 19a. *Imaging results for words per minute change and the left MOG.* A (Coronal View), B (Transverse View) and C (Sagittal View). MNI coordinates (x, y, z): -32, -75, 3. *t* value: 5.95.

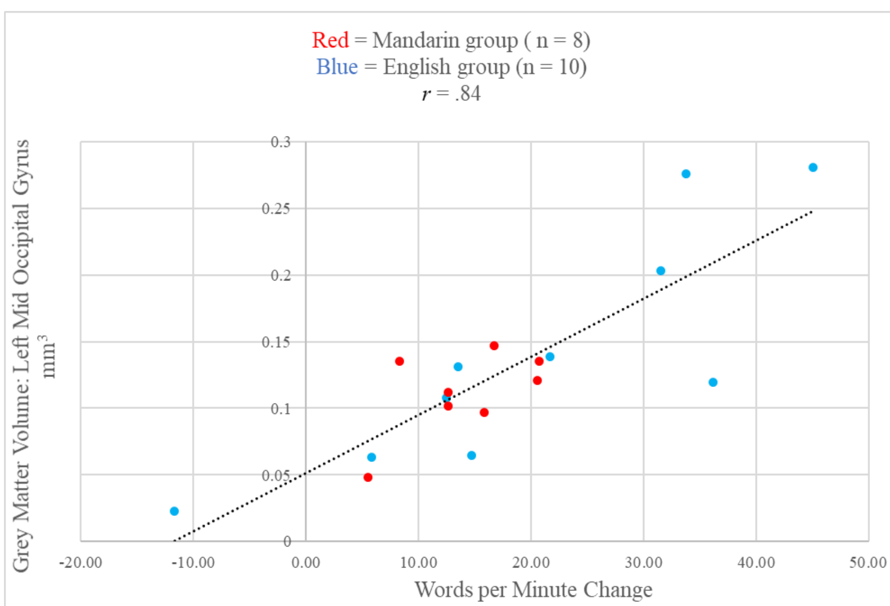


Figure 19b. *Correlation between pre-training left MOG GMV and words per minute change.* (r (16) = .84, p = <.001).

6 Discussion

1) Acoustic Analysis

The acoustic analyses demonstrated that both the English-L1 and Mandarin-L1 groups improved slightly on VOT and /i/ vowel production over French training. Specifically, average VOT durations became shorter and average /i/ formant frequencies became more similar to typical French values. The result that /i/ production provided the greatest indication of improvement among the vowels makes sense because /i/ occupies an extreme position (high and forward) in the vowel space. This characteristic usually allows for more variability in speech sounds that remain perceptibly as /i/. As such, /i/ is typically identified with a high rate of accuracy (Frieda et al. 2000), which is perhaps conducive to learning effects.

The Mandarin-L1 group improved more consistently on VOT than the English-L1 group, across both voiced and voiceless categories, and on more individual vowels. Although the group differences were not particularly substantial, this finding does contradict the prediction that the English-L1 group would improve more on the acoustic measures because Mandarin is more phonologically dissimilar to French than English is to French. Instead, this result may suggest that previous bilingual experience facilitated acoustic learning in the Mandarin-L1 group, in line with previous studies suggesting a bilingual advantage in phonetic acquisition (Spinu et al., 2018). However, the L1 groups were matched at pre-training on the Hindi Discrimination task, which assesses non-native phonological discrimination, raising questions concerning an interpretation of a bilingual advantage through this mechanism.

Alternatively, since Mandarin is relatively more phonologically dissimilar to French than English is to French, the Mandarin-L1 group may have improved more than the English-L1 group in French acoustic measures over a short period of time because the participants had more room for improvement. As such, the Mandarin-L1 group may have demonstrated a greater rate of initial learning changes because their early changes were more dramatic and obvious than the English-L1 group. Therefore, it is possible that these comparisons of improvement after 12 weeks of learning may not necessarily reflect ultimate group differences, for example after a year of learning. Future studies could use longer language training time scales to investigate dynamic rates of acoustic learning across participants of different language backgrounds.

Despite the positive findings of our acoustic analysis, there are a number of factors that may limit potential interpretations and are important to consider. First, we sampled speech recordings across tasks involving reading, repeating sentences, and free description, which all involve slightly different neural processes. These samples would have ideally been analysed separately if there were more data available, which is a potential avenue for future studies.

A second potential issue is that our reference vowels (which we used to quantify how French-like our participants' vowel production was) used European French as opposed to Quebecois French formant frequency values. This decision was made due to the absence of a reliable source for normal Quebecois French values. In addition, the instructors of the French training course showed a range of French accents, meaning Quebecois pronunciation was not the definitive learning reference for our participants. Nevertheless, due to the location of the present study in Montreal, it is likely that the participants were exposed to more Quebecois than European French in their everyday lives, so the quantification of their vowel pronunciation improvement may have been slightly mis-represented in the current analysis.

Finally, it is also important to note that there were inconsistent numbers of exemplars of many of the vowels produced, making it difficult to find consistent changes in the formant frequencies. In addition, due to the nature of the recordings, only a subset of French vowels was available for measurement, which might also have influenced the outcomes of the analyses. Therefore, future studies should aim to incorporate a greater range of vowels, with equal exemplars of each, to ensure balanced and comprehensive investigation of potential changes.

To conclude, despite a number of potential methodological issues, our analyses demonstrated that acoustic features of speech are a useful tool with which to index short-term articulation change, as they provide quantitative measures that are malleable during learning. In contrast, the results of our native accent ratings for the same speech samples were inconsistent, with poor consistency between the raters, and did not provide the same detail of improvement given the broad rating system.

2) Neuroplasticity

We did not observe any structural neuroplasticity in GMV after French training in either the whole-brain analysis or the subsequent voxel-specified analysis. The voxel-specified analysis was motivated by the idea that the brain regions where pre-training GMV correlated with behavioural improvement may be most likely to show neuroplasticity effects as they are presumably particularly involved in French articulation. Therefore, the lack of positive results even after this more sensitive analysis strongly suggests a lack of GMV longitudinal changes in this sample.

There are a number of reasons for the lack of neuroplasticity effects. A possibility is that the intensive French training simply may not have been effective enough to elicit neural change. This may be the case because our behavioural analysis did not consistently show a significant improvement in French proficiency.

In addition, factors inherent to the study design may have reduced the likelihood of neuroplasticity. First, the time scale of 12 weeks may have been too short to observe structural changes in GMV (Elmer et al., 2014). Previous studies into neuroplasticity after language learning have more commonly reported functional neuroplasticity effects, for example in IPL activation after 12 weeks of French training (Barbeau et al., 2016). Furthermore, changes in functional connectivity have emerged even after short-term (single session) speech category training. For example, Feng, Yi, and Chandrasekaran (2018) tasked native English speakers to learn to categorise Mandarin lexical tones and collected DTI data over training. Results showed that, with learning, functional coupling between the putamen and left superior temporal gyrus (STG) increased during error processing, implying a critical and modifiable role of auditory cortico-striatal circuitry in mediating the acquisition of new speech categories.

In contrast, structural plasticity, for example in GMV and cortical thickness in language-relevant areas, has generally occurred over longer time frames (Stein et al., 2012; Legault et al., 2019). Therefore, future studies aiming to explore potential structural brain changes after language training should incorporate longer time scales into the study design. This approach also enables exploring potential dynamic changes in brain structure as individuals may switch between different learning strategies as their proficiency develops. For example, switching between reflexive and reflective learning in acquiring non-native speech categories may involve different brain regions that behave competitively, thereby inducing antagonistic changes in brain structure over time.

Lastly, structural brain changes can also depend on age of acquisition (AoA). Specifically, earlier language learning can induce increased anatomical changes (Penicaud et al., 2013). Therefore, the lack of plasticity in our participants is perhaps unsurprising due to their relatively high AoA (mean age: 22 years). It is possible that neuroplasticity effects may have been observed in a similar context but with adolescent rather than adult participants (Yamamoto and Sakai, 2017). As such, future studies could perhaps compare how language background and age of learning may interact in producing neuroplasticity effects after language training.

3) Biomarkers

The biomarker analyses were designed to capture the extent to which the variance in French articulation improvement could be explained by pre-training brain structure. We pinpointed various pre-training differences in regional GMV across our participants that correlated with learning success in different behavioral tasks. GMV differences occurred in subcortical regions important for implicit motor processing, regions important for phonological processing, and in cortical areas that may relate to the specific cognitive demands of reading or sentence repetition.

First, considering the biomarkers for our acoustic measures, for VOT, improvement was predicted by higher GMV in the right cerebellum (lobule VIIb/VIIIb). For /i/ production, improvement was comparably predicted by higher pre-training GMV in the left and right cerebellum (lobule VIIb/VIIIb in both hemispheres). Acquiring new acoustic patterns is a subconscious process that may rely on implicit motor and sensorimotor learning and accurate phonological processing. Our findings make sense because the cerebellum has been linked to both of these functions. Specifically, the bilateral cerebellum is important for motor adaptation required in implicit motor skill learning (Doyon, Penhune, and Ungerleider, 2003), perhaps through its putative role in automatic error correction through generating sensory error during movement (Grafton et al. 2008). Therefore, the cerebellum may underlie the timing and fine motor coordination of articulators during VOT and vowel production, and thus may be important for acoustic and articulatory learning. Concerning phonological processing, the right cerebellum (lobule VIIb) has also been implicated in cerebrocerebellar phonological working memory circuits in concert with the IPL (BA40) (Chen and Desmond, 2005). Therefore, through this mechanism the cerebellum may be important for acquiring and adjusting accurate speech sounds, as in learning the acoustic features of speech.

An additional finding was that VOT improvement was predicted by lower GMV in the left dorsal and ventral PMC and the left dmPFC. In contrast to the cerebellum, the PMC and dmPFC are important for developing explicit knowledge of practiced motor sequences (Honda et al., 1998; Robertson, 2009) and more explicit (reflective) processing in speech category learning (Yi et al., 2014). As such, greater GMV in areas underlying more explicit motor learning may actually interfere with the implicit motor learning that is perhaps necessary for learning new acoustic patterns successfully. However, it is beyond the scope of the current study to establish whether a greater "reliance" on explicit over implicit motor processing may have interfered with acoustic learning. Future studies could perhaps use functional, task-based measures and protocols designed to establish competition between implicit and explicit motor learning mechanisms to explore such a question.

Next, considering Articulation Rate, improvement was predicted by higher pre-training GMV in the left caudate and the right IPL. This measure was meant to provide an indicator of speech fluency, which may rely on both motor control and phonological processing similarly to the previous acoustic measures. The left caudate has been linked to both of these functions in relation to speech production (Grogan et al., 2012; Abutabeli and Green 2007). In particular, the caudate is implicated in implicit motor processing (Doyon et al., 2009) such as implicit sequence learning (Destrebecqz et al., 2005) and cortico-striatal circuits important for implicit (versus explicit) speech category learning (Yi et al. 2014). In addition, the caudate is important in phonemic fluency tasks (Grogan et al. 2009) and in detecting phonological anomalies (Tettamanti et al. 2005b), perhaps through its role in procedural reinforcement learning (Tricomi et al. 2006). The IPL has also been linked functionally to phonological working memory (Koelsch et al. 2009; Chen and Desmond, 2005) and speech category learning (Yi et al. 2014). As such, previous research has already pinpointed the bilateral IPL as a structural biomarker for accurate pronunciation of non-native phonemes (Golestani and Pallier, 2006). Therefore, this set of results implies that GMV in brain regions important for implicit motor and phonological learning facilitates faster paced free speech generation.

Next, considering Sentence Repetition, accuracy improvement was predicted by higher pre-training GMV in the left MTG and the left globus pallidus. This finding is consistent with the Articulation Rate result, as in both cases the left basal ganglia, important for implicit motor processing, was relevant to improvement in both measures of articulation. The globus pallidus is the main output nucleus of the basal ganglia and involved in the implicit regulation of voluntary movement (Gillies et al. 2017). In support, lesioning the globus pallidus is associated with impairment of new motor skill acquisition but not the retention of already-learned skills (Desmurget and Turner, 2008). Therefore, this structure may contribute to L2 sentence repetition accuracy due to the task's demands for acquiring new motor speech patterns. In addition, the MTG result fits in well with models conveying a key role in semantic processing for the MTG in auditory sentence processing circuits (Friederici, 2002). Therefore, this result demonstrates that biomarkers for language learning may be associated with the specific cognitive demands of the task in question.

Finally, considering Words per Minute, improvement was predicted by higher pre-training GMV in the left MOG, slightly posterior to classic VWFA coordinates in the left ventral occipitotemporal cortex, which is thought to respond specifically to visual orthographic stimuli (Cohen and Dehaene, 2004). This finding replicates previous studies implicating similar brain regions in the acquisition of literacy (Dehaene et al., 2010) and conversely in the neurocognitive basis of dyslexia (Paulesu et al., 2001). Therefore, similar to the previous finding linking the MTG to sentence repetition accuracy improvement, this result demonstrates that biomarkers for language learning depend on the specific cognitive demands of the task in question.

In terms of interpreting our biomarker results as a whole, our most consistent finding was that GMV in areas important for implicit motor learning and phonological processing is predictive of subsequent articulation improvement across different speech production tasks. This finding relates to previous literature which argues that domain-general cognitive and learning abilities underlie language aptitude (Frankish, 2010). For example, studies have advocated for the particular importance of working memory (Robinson, 2005) and procedural learning (Linck et al. 2014) in predicting L2 proficiency. Although the current study did not explicitly link working memory or procedural learning to brain structure, we did report biomarker effects for behaviours that arguably require these abilities in brain regions that arguably underlie them. Other studies have made similar links between brain structure and behavior in language learning (Golestani et al., 2002). Overall, it is an important goal of future studies to explore how proposed structural biomarkers for language aptitude may actually relate to subsequent behavioural improvement, including using both functional and structural data to clarify the possible mediating effects of brain function and corresponding cognitive abilities.

However, there are a number of outstanding questions regarding the biomarker analysis that may limit our potential interpretations. First, it remains unclear why GMV in different brain regions important for both implicit motor and phonological learning would correlate only with specific measures of articulation. For example, why did the caudate not emerge as a biomarker for the acoustic measures as well as for Articulation Rate, and vice versa for the cerebellum, if both regions are important for pronunciation. Similarly, why did the globus pallidus, important for new motor skill acquisition, emerge as a biomarker only for sentence repetition accuracy when all of the tasks are designed to capture articulatory learning and presumably require motor skill learning. For this particular question, perhaps the nature of the sentence repetition task demanded more immediate motor skill acquisition, whereas the acoustic and other articulatory measures tapped more into the retention of already-learned motor skills, acquired over the 12 weeks training, for which the globus pallidus may be less important (Desmurget and Turner, 2008). In any case, future studies could focus on exploring the connectivity of networks involving these brain regions to help to resolve such inconsistencies, through techniques like resting-state fMRI, as opposed to simply measuring brain structure in different regions in isolation.

Finally, a second potential issue with the biomarker analysis is the certainty of our interpretation that pre-training differences in GMV reflect the structural neural basis of language aptitude in our participants. It was beyond the scope of the current study to disentangle how experience (e.g. the amount and type of environmental input) interacts with the initial status of relevant brain networks that influence learning, and the extent to which individual differences in predisposition are themselves the outcome of plasticity due to previous experiences in other domains and/or (epi)genetic variability. Researchers have argued that pre-existing individual differences should assume a greater importance in the literature (Kanai and Rees, 2011), and future studies could explore how the neural basis of language aptitude may incorporate the capacity for learning and neuroplasticity, as opposed to just static structural measures in language-relevant brain regions.

4) Language Background

Aside from the group differences in the acoustic analyses addressed above, we did not detect consistent behavioral group differences in French improvement. This finding counteracts the prediction that the English-L1 group may find it easier to learn French than the Mandarin-L1 group, given the closer typological distance between their native and target language. Instead, the different hypothesised rates of French improvement between the groups may have been confounded by uncontrolled factors related to the characteristics of individual participants, which had an exaggerated effect on our results due to the small sample size. For example, variation in personality traits such as conscientiousness can affect language learning (Molaei, 2016). Therefore, future studies should aim to use larger sample sizes to rule out potentially confounding group differences beyond cognitive and intelligence measures.

Moreover, there are a number of reasons for the lack of significant group differences in French improvement based on a distinction between monolingualism (in the English-L1 group) and previous bilingual experience (in the Mandarin-L1 group). Most obviously, our groups were matched at pre-training on working memory, cognitive control, and non-native phonological discrimination. Therefore, in this sample, there is no basis to suggest that existing bilingualism in the Mandarin-L1 group facilitated French learning through these mechanisms, as previous research has supported (Bartolotti and Marian, 2012; Kaushanskaya and Marian, 2009).

In addition, our groups were not entirely homogenous in terms of language background. For example, the Mandarin-L1 group showed a great deal of variation in their reports of English use in daily conversations (between 20-70%). In addition, the English-L1 group was only relatively consistently monolingual, with 3/10 reporting their % English use in conversations as only between 75-85%. To acknowledge this variability, we considered the groups in terms of differences in language background in-stead of monolingual and bilingual as categorical variables. However, there is certainly still a spectrum of experiences within our groups that might have differentially affected brain structure and function (DeLuca, Rothman, et al., 2019). For example, experience-based factors such as the amount of L2 use in social settings and the nature of L2 input may have varied across our Mandarin-L1 participants' use of English and yielded specific adaptations in the brain that influenced our results (Li et al., 2014). Therefore, perhaps with more homogeneous groups, our behavioral results would have revealed a clear-cut group difference in French learning.

Nevertheless, despite these concerns, the main focus of the group comparison was the effect of native language (Mandarin or English) on French learning, rather than the effect of previous bilingual experience. Native language was an objective and consistent factor between and within the groups and provides a valid basis for comparison despite other potential sources of heterogeneity.

In terms of our biomarker analysis, nearly all of our detected biomarkers predicted French articulation improvement across both of our groups, independent of existing bilingualism and native language. This conclusion is largely consistent with the idea that language aptitude is based in part on pre-existing cognitive, motor, and perceptual abilities. These abilities may be influenced themselves by language background, among many other experiential factors, but are not necessarily dictated by it. Future studies could aim to disentangle whether such abilities operate in a domain-general manner beyond language, or, alternatively, if they could be specific to language. In addition, although the present study suggests that language background may not necessarily influence structural biomarkers for French learning, future studies could also explore whether biomarkers for success in language acquisition may depend on the target language in question and its particular demands and characteristics. For example, Qi, Han, and Garel (2014) reported that right hemisphere white matter structure may be more important than left hemisphere structure for superior learning of Mandarin compared to other languages without the same tonal and visuospatial properties, which are typically processed in the right hemisphere.

Nevertheless, despite the general consensus, we reported one group difference in our VBM results; that GMV in the left dorsal ACC predicted Articulation Rate change in the Mandarin-L1 group but not the English-L1 group. This finding is interesting because it relates to previous research that indicates the ACC is 'tuned' for bilingualism through a role in conflict monitoring (Abutalebi et al. 2011). Therefore, the ACC may be more important for participants who were learning an L3 on top of a developing L2, as opposed to just an L2, for managing their multiple language systems. However, the lack of consistency of this result across our other behavioral measures warrants caution in interpretation. In future studies, larger sample sizes and more homogenous groups would be required to establish the ACC as a part of the neural basis of language aptitude only for individuals with bilingual experience.

Finally, the group comparisons regarding our biomarker findings were only driven by late, sequential bilingual participants, and may not necessarily apply to other bilingual individuals who differ in AoA. Specifically, late bilingual systems may be less lateralised than early bilingual systems, display a high degree of variability between individuals, and be also more likely to include activation of control regions such as the ACC when speaking the L2 but not the native language (Dehaene et al. 1997). Therefore, future studies should explore how the neural basis of language aptitude may vary within bilingual individuals depending on how early they acquired their L2.

7 Conclusion

7.1 Conclusions

The present study yielded 4 main conclusions. First, participants remained stable between pre and post-training on the untrained language (English), but improved on a number of French behavioral measures, including the acoustic measures of VOT and /i/ vowel production. Second, in this sample, 12 weeks of French training in adulthood was insufficient to elicit structural effects of neuroplasticity in GMV. Third, there were different GMV biomarkers for different measures of French articulation. Biomarkers occurred in brain regions important for implicit motor processing (the bilateral cerebellum and the left basal ganglia), phonological processing (the IPL and the left caudate), and visual and auditory areas related to the requirements of the different behavioral tasks. In contrast, brain structure in regions underlying explicit motor processing actually hindered learning new acoustic patterns. Finally, there was a slight effect of language background on acoustic learning and one of the biomarker results in the ACC. However, most biomarkers were independent of language background.

Overall, this research provides insight into why individuals may vary in their ability to learn a language, in this case French, in adulthood, and how we can accurately index such learning with more quantitative and fine-grained approaches. Our results demonstrate that the neural basis of language aptitude for articulatory learning includes brain regions important for implicit motor learning and phonological processing, but also can differ according to the specific behavioral task in question. Although language background partially influences the acquisition of accurate articulation in a target language in adulthood, overall, more general motor, cognitive, and perceptual abilities seem to be more important for predicting success.

7.2 Limitations

Our study design was limited in a number of ways. First, our sample size was only 18, meaning there is not a high level of statistical power to draw large-scale conclusions. Second, we cannot access information on the effectiveness of teaching and learning in the intensive French training program. We must assume that the course was equally attended and attempted across our individuals, meaning any differences in outcome reflect variation in language aptitude in stead of motivation or effort. Finally, when the present study was originally conceived, the acoustic analyses that we eventually conducted were not planned, meaning the quality of recordings was not entirely conducive to fine-grained phonetic analyses.

7.3 Future Directions

Future studies in this area should aim to tackle a number of outstanding questions, incorporating larger sample sizes, a greater mix of language backgrounds to systematically compare the effects of native language and bilingualism, and more precise recording techniques. In addition, future training studies should aim to include training schedules with longer time frames, which would increase the likelihood of neuroplasticity effects, and also shed light on potential changes in neural compensation as learners switch between reflective and reflexive learning systems as their proficiency develops. Moreover, there is great potential in using acoustic approaches to objectively quantify speech articulation. This potential could help be realised by machine learning technology, which could be used with great effect to analyse the large amount of speech sample data that is required to extract accurate VOT and vowel formant measurements. Finally, since domain-general motor and cognitive abilities may contribute to language aptitude, a future step is considering learning interventions to target and improve such factors in the hope of transferring these skills to facilitate language learning.

Appendix I

Language Experience and Proficiency Questionnaire (LEAP-Q)

Health and Language History Questionnaire

"The purpose of the following questionnaire is to obtain more information about your language and health history for the purpose of matching the groups included in the current study on bilingualism and language learning. If you are uncomfortable answering any of the questions, feel free to leave them blank. Thank you!"

Section 1: Demographic Information

1. Date of Birth (day/month/year):

2. Age:

3. Sex: Choose Male/Female

4. Handedness: Choose: Left/Right/Both

5. Education: What is the highest level of education that you have completed? You can include information such as "attended but did not complete"

   Primary school

   High School; where did you completed high school (province, Country)?

   CEGEP

   College/University undergraduate degree (e.g., BA, BSc)

   Graduate degree (e.g., Master's degree)

   Graduate degree (e.g., Ph.D., MD)

   Other; please specify

6. What is your current marital status?

   Single – never married

   Married / Common-law

   Separated

   Divorced

   Widowed

   Cohabit

7. What is your main occupation?

8. If you are married, what is your spouse's highest level of education and their main occupation?

9. What is your mother's highest level of education and her main occupation (if retired, what was her occupation prior to retirement?)

10.What is your father's highest level of education and his main occupation (if retired, what was his occupation prior to retirement?)

11.Where were you born? If not in Canada, how long have you been in Canada?

12.Where were your parents born? (If not in Canada, please indicate if they are currently in Canada, how many years they have been in Canada, their native language and other languages that they speak):

13.Do you play a musical instrument? Choose: Yes / Yes, but not well / No

If "yes"

a. Do you have any formal musical training? Choose; Yes, from when I was a child / Yes, in primary school / Yes, in high school / Yes, in primary and high school / Yes, I currently take lessons / No

b. Do you still play? Choose: Yes / No

c. How frequently? Choose: Less than once a month / Once a month / 2-3 times per month Once per week / 2-3 times per week / Every day

d. Can you read music? Choose: Yes / No

e. Do you consider yourself a musician? Choose: Yes / No

Section 2: Language Background and Experience

1. Do you speak more than one language? Choose: Yes / No

   If you answered "no", skip to the next section

   If you answered "yes", please list the languages that you speak in order of fluency (with the most fluent first):

2. Please rate your current ability on reading, writing, speaking, and listening for all languages you know according to the following scale (1 – Very Poor, 2 – Poor, 3 – Fair, 4 – Functional, 5 – Good, 6 – Very Good, 7 – Native-like):

3. Have you ever taken a standardized language proficiency test in your non-native language(s) (e.g., TOEFL? If yes, please indicate the name of the test, the languages assessed, and the scores that you received. If you can't remember, please guess. If you remember only the percentile of your score, write it in the place of the score.)

4. Do you have a foreign accent in the languages that you speak? Please rate how strong you think your accent is according to the following scale (1 – None, 2 – Little, 3- Some, 4 – Intermediate, 5 – Strong, 6 – Very Strong, 7 – Extremely Strong):

5. At what age did you first start to learn each language in terms of speaking (at what age did you speak your first words?), reading, and writing, and the number of years you have spent learning each language (cumulative).

6. Please indicate the age at which you started to learn each language in the following situations – indicate the age in the boxes for only situations that are relevant (at home, at school, after immigrating to the country where spoken, informal setting (e.g. nannies or friends), software (e.g. Rosetta Stone), Other (please specify):

7. Please indicate the language(s) used by your teachers for general instruction (e.g., history, math, science) at each schooling level. If you switched language within a level please indicate the level and the languages.

Primary School:

High School:

CEGEP:

College/University:

Other:

8. Have you ever lived or travelled in another country for more than three months where you were required to speak another language other than your native language(s)? If so, please indicate the country, your length of stay and the year that you visited, the language(s) that you learned or tried to learn, and your frequency of use of the language while visiting the country and currently. Please use the following scale (1 – Never, 2 – Rarely, 3 – Occasionally, 4 – Sometimes, 5 – Frequently, 6 – Very Frequently, 7 – Always):

9. How good do you think you are at learning new languages (e.g., relative to friends or people you know). Circle one (Very poor, Poor, Fair, Neutral, Good, Very good, Excellent):

10.At what age do you consider that you became fluent in each language in terms of speaking, reading and writing? Please indicate an age in each box; if you do not consider yourself fluent please indicate “not fluent”.

11.Please estimate the total number of hours each day that you spend engaged in the following activities, and indicate what percentage of that time you spend engaged in that activity in each of the languages that you know (please write down the languages). If you are not currently engaged in an activity using that language write “0”; the total percentage for each activity should equal 100%.

Listen to radio / watching TV

Reading for fun

Reading for work

Reading on the internet

Writing emails to friends

Writing articles / papers

Other (specify):

12.Please estimate the percent of conversations that take place in each of your languages, and what percentage of that is with the following people. The total across languages should equal 100% and the total within each language should equal 100%.

Language

Family members

Friends

Classmates

Co-workers

13.How often do you use your languages for the following activities? Use the following scale and fill in the number in the table (1 – Never, 2 – Rarely, 3 – Occasionally, 4 – Sometimes, 5 – Frequently, 6 – Very Frequently, 7 – Always):

Arithmetic (e.g., count, add)

Remember numbers (e.g., student ID, telephone)

Dream Think

Talk to yourself

Express anger or affection

14.What proportion of your current friends are speakers of the languages that you know well? Please indicate the language and the percentage of your total number of friends that speak that language (the total should equal 100%). Include a separate category for bilingual friends and please indicate the languages that you speak with them.

15.In which language (among your two best languages) do you feel you usually do better or feel more comfortable? Indicate the language for each condition.

At home

At work / At school

At a party or other social context

16.Do you mix words or sentences from two languages in your own speech (e.g., say a sentence in one language but use a word or phrase from another in the middle of the sentence)? Yes / No

If you answered "no", please move on to the next section.

If you answered "yes", please continue with the following questions

17. a) List the two or more languages that you mix with different people, and estimate the frequency of mixing/switching in normal conversation according to the following scale (1 – Never, 2 – Rarely, 3 – Occasionally, 4 – Sometimes, 5 – Frequently, 6 – Very Frequently, 7 – Always):

Family members

Friends

Classmates

Co-workers

17. b) Under what situations from those listed below are you most likely to mix/switch between two languages, and which languages do you mix/switch between? Please list all language combinations that apply to each situation (e.g., English and French; from English to French).

When I don't know the word in one language

A word comes to me faster in one language

It is difficult for me to control which language I am speaking in

I switch between languages on purpose

Other (specify):

17. c) Please indicate if there are situations in which you are more likely to mix or switch between languages and what those situations are.

17. d) Please indicate if there are situations in which you think that it is inappropriate to mix or switch between languages, and what those situations are.

18.Do you feel that you are bilcultural or multicultural (e.g., growing with parents or relatives from different cultures, or you lived in different cultures for extended periods of time)? Yes / No

If "yes", which culture (and its language) do you identify more strongly with? Use the following examples and scale to indicate the strength of your cultural identification (1 – None, 2 – Very Weak, 3 – Weak, 4 – Intermediate, 5 – Strong, 6 – Very Strong, 7 – Extremely Strong):

Culture and its Language

Like its food

Like its music

Like its art

Like its cities and landmarks

Will root for its athletic teams

19.Is there anything else that you think is interesting or important about your language background or language use?

Section 3: Health Information

1. Do you have any visual problems (e.g., cataract, colour blindness, wear glasses)? Yes / No

2. Do you have any hearing problems (e.g., hearing loss, do you wear a hearing aid)? Yes / No

3. Have you ever had a head injury? Yes / No

If "yes",

What was the cause?

What was the outcome?

4. Do you have a history of neurological disorder? Yes / No

5. Have you ever had any major surgery? Yes / No

What for?

6. Do you have any metal prostheses, screws, plates or fragments? Yes / No

7. Do you have any piercings or tattoos? How many, and where are they located? Yes / No

8. Do you have any allergies? Yes / No

9. Are you claustrophobic? Yes / No

10.Are you pregnant? Yes / No

11.Have you ever had an MRI before? Yes / No

For what?

12.Do you currently take any medications? Yes / No

If "yes", please list the medications and indicate what condition you are taking them for and how long you have been taking them for Medication Reason for consumption Duration of consumption

13.Do you drink alcohol? Yes / No

If "yes", approximately how many drinks of alcohol do you have per week?

14.Do you use non-prescription drugs for recreational purposes?

If "yes", which ones and how many times per week?

Appendix II

Story Reading and Comprehension Questions

English

Lindsay decided she needed a cabinet for her new dishes. She measures the empty space in her kitchen before leaving for the lumberyard. After looking at the cedar and oak boards, she concluded that she much preferred pine. She went to the hardware store after buying her wood. She purchased brass handles and hinges, as well as a big hammer. The price was thirty-two dollars. She rushed to her workshop in the basement as soon as she got home. Lindsay wanted to finish her corner cabinet before the Christmas holidays.

Questions

1. What was the name of the person in this story?

2. What did she decide she needed?

3. What did she do before going out?

4. Where did she go first?

5. What types of wood did she look at?

6. Which did she like best?

7. Where else did she go?

8. What did she buy while she was there?

9. Where did she go after leaving that place?

10. Is the person an adult or a child?

11. Is the person a man or a woman?

12. Where was she going to put the cabinet?

13. Where did she buy the wood?

14. What kind of hinges was she going to put on the cabinet?

15. Where was she going to build her cabinet?

16. When (at what time of year – what season) did Lindsay build her cabinet?

Français

La chambre d’invités de Michelle paraissait défraichie. Elle a décidé de redécorer avant la visite de sa sœur, prévue pour Pâques. Avant de conduire jusqu’au centre d’achats, elle a trouvé un échantillon de son papier-peint et l’a emporté avec elle. Après avoir contemplé les peintures orange et écarlate, Michelle prit vert pomme pour un mur et beige pour le plafond. Elle a aussi choisi un pinceau, un petit rouleau et de la térébenthine, qui ont coûté vingt-quatre dollars en tout. Elle s’est arrêtée chez un fleuriste sur son chemin et a commandé une grande plante pour être livrée le jour même.

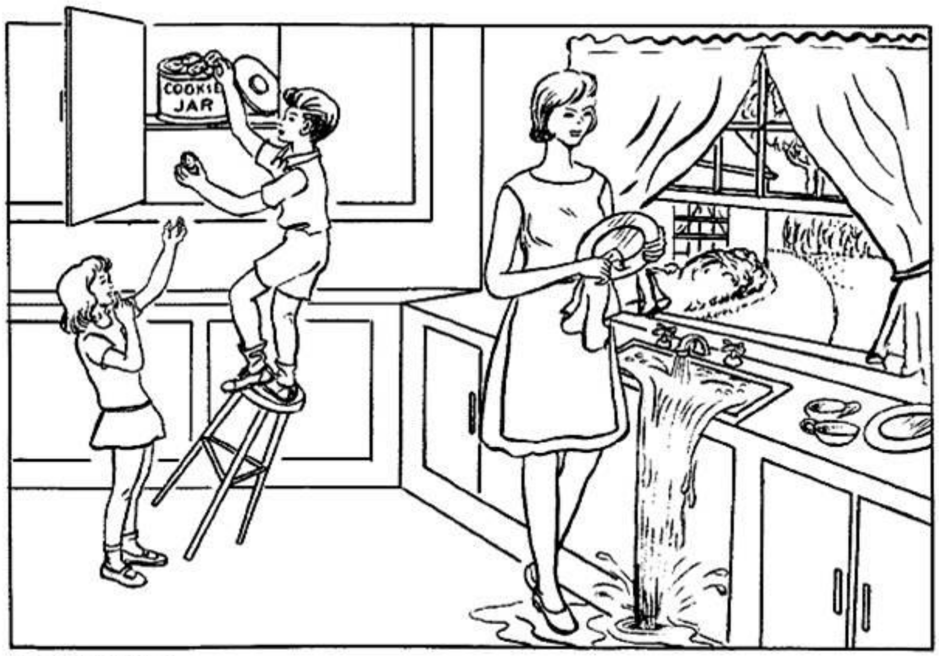
Questions

1. Quel était le nom de la personne de cette histoire?
2. Que voulait faire cette personne?
3. Quand voulait-elle avoir fini?
4. Qu'a-t-elle fait avant de sortir?
5. Où est-elle allée?
6. Quelles couleurs a-t-elle regardées?
7. Quelles couleurs a-t-elle achetées?
8. Qu'a-t-elle acheté d'autre?
9. Où est-elle allée en dernier?
10. Qu'est-ce qu'elle a acheté là-bas?
11. Est-ce que la personne dans cette histoire est adulte ou enfant?
12. Est-ce que cette personne est un homme ou une femme?
13. Où sa sœur allait-elle rester?
14. Comment s'est-elle rendue au centre d'achats?
15. Est-ce que cette personne a acheté sa peinture avant ou après avoir pris son papier-peint?
16. Est-ce qu'elle a arrosé la plante en rentrant chez elle?
17. A quel moment de l'année (quelle saison) a-t-elle redécoré la pièce?

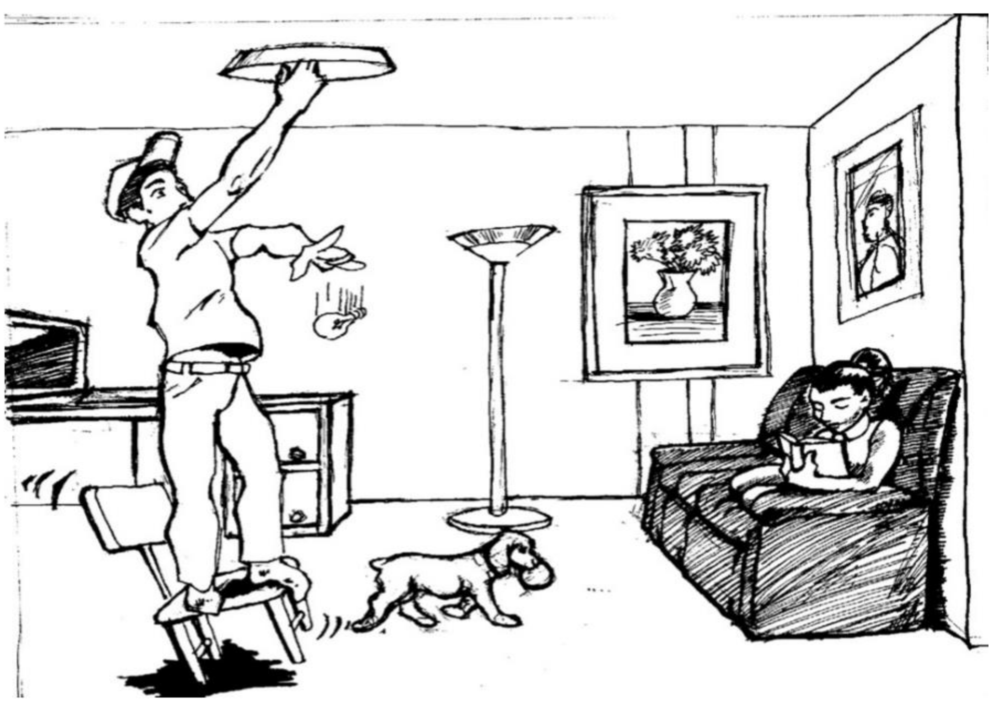
Appendix III

Picture Description Task Stimuli

English – Cookie Theft Picture



French – Lightbulb Picture



Appendix IV

Sentence Repetition Stimuli

English

1. Does anyone know who the new teacher is?

2. The kindergartener cannot cross the street by himself.

3. The play castle was built by the girls and boys.

4. Because tomorrow is Saturday, we can stay up late tonight.

5. The book was not returned to the library by the teacher.

6. The coach could not find the uniforms that the team wore last year.

7. The girl stopped to buy some milk, even though she was late for class.

8. My mother is the nurse who works in the community clinic.

9. The boy bought a book for his friend who likes short stories.

10. If the rain doesn't stop before noon, the field trip will have to be canceled.

11. The computers and printers were donated by the school board.

12. The student who won the award at the art show was very excited.

13. The class that sells the most tickets to the dance will win a prize.

14. After the students had finished the book, the teacher asked them to write a report.

15. Coach gave the trophy to the team that won the track meet on Saturday.

16. The students collected and repaired the toys, and sold them at the fair.

17. Today we must have lunch early, go to the library, and finish our art projects.

18. When the students finished studying, they decided to get something to eat before going home.

19. The librarian has twelve new eight-grade science books reserved for us.

20. If we had gone straight home after the game, we would not have missed our curfew.

21. Before they walked across the stage for graduation, the students lined up in alphabetical order.

22. If I don't have to work this weekend, I should be able to finish my research paper for English.

23. Before the students were dismissed for lunch, they were told by the teacher to turn in their assignments.

24. The math teacher sorted, labeled, boxed, and delivered the calculators.

French

1. Est-ce que quelqu'un sait qui est le nouveau professeur?

2. L’élève de maternelle ne peut traverser la rue seul.

3. Le château de sable a été construit par les filles at les garçons.

4. Parce que demain c’est samedi, nous pouvons veiller tard ce soir.

5. Le livre n’a pas été retourné à la bibliothèque par le professeur.

6. L’entraineur n’a pu trouver les uniformes que l’équipe portait l’an dernier.

7. La fille s’est arrêtée pour acheter du lait, même si elle était en retard pour l’école.

8. Ma mère est l’infirmière qui travaille à la clinique communautaire.

9. Le garçon a acheté un livre pour son ami qui aime les romans.

10. Si la pluie n’arrête pas avant midi, la sortie devra être annulée.

11. Les ordinateurs et les imprimantes ont été donnés par la commission scolaire.

12. L’élève qui a gagné le prix du spectacle d’art était très heureux.

13. La classe qui vend le plus de billets pour la danse gagnera un prix.

14. Après que les élèves ont eu terminé le livre, le professeur leur a demandé d’écrire un rapport.

15. L’entraineur a donné le trophée à l’équipe qui a gagné la compétition d’athlétisme de samedi.

16. Les élèves ont recueilli puis réparé les jouets et les ont ensuite vendus au marché aux puces.

17. Aujourd’hui, nous devons manger tôt, aller à la bibliothèque et finir nos projets d’arts plastiques.

18. Lorsque les élèves ont eu terminé d’étudier, ils ont décidé de manger quelque chose avant d’aller à la maison.

19. Le bibliothécaire a réservé douze (12) nouveaux livres de sciences de niveau secondaire pour nous.

20. Si nous avions été á la maison après la partie, nous n’aurions pas manqué notre couvre-feu.

21. Avant de traverser la scène pour la graduation, les étudiants étaient placés en ordre alphabétique.

22. Si je n’ai pas à travailler en fin de semaine, je devrais être capable de compléter mon travail de recherche en anglais.

23. Avant que les étudiants reçoivent la permission d’aller diner, ils se sont fait dire par leur professeur de remettre leur devoir.

24. Le professeur de mathématiques a trié, étiqueté, puis distribué les calculatrices.

SUPPLEMENTARY MATERIALS

I. The Simon Task

The participants took an arrows version of the Simon Task. There were 6 trial blocks (2 blocks for each of the 3 conditions, presented in a counter-balanced order). The 3 conditions were “control”, “reverse”, and “conflict”. In total there were 96 trials of each type for a total of 384 trials. In the control condition participants had to indicate in which direction centrally presented arrows were pointing (either left or right using left and right response keys), with their response times as the outcome measure. In the reverse condition participants had to indicate the opposite direction to which the arrow was pointing (e.g. left key response for a rightward pointing arrow). In the conflict condition participants had to indicate the direction of the arrow using the left and right response keys, but trials were randomly intermixed to be either congruent (where the directional arrow was presented on the same side of the laptop screen as the correct response) or incongruent (where the arrow was presented on the opposite side of the screen to the correct response). So, participants had to ignore the irrelevant spatial information from the position of the stimulus in order to respond to the direction of the arrow.

The 3 conditions allowed us to calculate 3 different components of the Simon effect that were used in the present study. These were; (1) the Simon effect (the increase in response time for incongruent relative to congruent trials within the conflict condition); (2) response inhibition (the ability to inhibit a habitual response, calculated as the increase in response time for the reverse compared to the control condition), and; (3) interference suppression (the ability to suppress interfering spatial information; calculated as the increase in response time for the conflict compared to the control condition).

II. Voice Onset Time Analysis

In English, there were no significant differences between the L1 groups for any voiceless or voiced plosives at both time points (all p's > .376). In French, there were no significant differences between the groups for any voiceless plosives at either time point, but there was for the voiced token /b/ at pre-training (t (16) = 2.36, p = .031) where the English-L1 group produced significantly shorter tokens. There were no significant differences between the groups in the extent of French VOT change for either voiced or voiceless tokens (all p's >.169).

French VOT duration was typically shorter than English VOT duration. We ran a univariate 3-way ANOVA with Place of Articulation (PoA) (3 levels: labial, dental, or velar), Voicing (2 levels: voiced or voiceless), and Language (2 levels: French or English) as factors. Results showed significant main effects for PoA (F (2, 312) = 120.16, p = <.001, $\eta p^2 = 0.435$), Voicing (F (1, 312) = 1682.08, p = <.001, $\eta p^2 = 0.844$) and Language (F (1, 312) = 223.95, p = <.001, $\eta p^2 = 0.418$) , significant 2-way interactions for Language * Voicing (F (1, 312) = 33.89, p = <.001, $\eta p^2 = 0.098$), Language * PoA (F (2, 312) = 140.29, p = <.001, $\eta p^2 = 0.473$) and Voicing * PoA (F (2, 312) = 33.21, p = <.001, $\eta p^2 = 0.176$), and a significant 3-way interaction (F (2, 312) = 79.45, p = <.001, $\eta p^2 = 0.337$).

Post-hoc paired sample t-tests (Bonferroni corrected) showed a significant difference between French and English voiced tokens at pre-training (t (17) = 9.09, p = <.001) and at post-training (t (17) = 9.11, p <.001), and between French and English voiceless tokens at pre-training (t (17) = 4.35, p = <.001) and at post-training (t (17) = 6.57, p = <.001). For each of these 4 comparisons, the average VOT duration was significantly longer for English than French.

Table 1. Voiced VOT Durations

	French						English					
Time	Pre-Training			Post-Training			Pre-Training			Post-Training		
Plosive	b	d	g	b	d	g	b	d	g	b	d	g
n	16	72	12	16	72	12	61	37	22	61	37	22
English-L1	18 *(4)*	21 *(3)*	19 *(3)*	18 *(2)*	21 *(3)*	16 *(4)*	26 *(4)*	23 *(5)*	53 *(3)*	25 *(4)*	22 *(2)*	52 *(4)*
Mandarin-L1	23 *(5)*	22 *(3)*	17 *(3)*	19 (3)	20 *(4)*	17 *(4)*	26 *(4)*	22 *(4)*	52 *(5)*	27 *(6)*	24 *(5)*	54 *(6)*

The average voiced VOT values (in msec) at pre and post training in English and French for the L1 groups separately. Mean *(SD)*.

Table 2. Voiceless VOT Durations

	French						English					
Time	Pre-Training			Post-Training			Pre-Training			Post-Training		
Plosive	p	t	k	p	t	k	p	t	k	p	t	k
n	86	91	59	86	91	59	22	42	34	22	42	34
English-L1	53 *(5)*	54 *(8)*	56 *(8)*	51 *(6)*	47 *(7)*	53 *(8)*	39 *(8)*	69 *(4)*	71 *(4)*	37 *(7)*	67 *(3)*	68 *(4)*
Mandarin-L1	58 *(8)*	52 *(8)*	54 *(4)*	55 *(6)*	45 *(9)*	51 *(4)*	40 *(6)*	70 *(7)*	70 *(6)*	43 *(17)*	70 *(7)*	70 *(6)*

The average voiceless VOT values (in msec) at pre and post-training in English and French for the L1 groups separately. Mean *(SD)*.

III. Vowel Formant Analysis

Table 3. French Vowel Formant Frequencies

		Pre-Training English-L1		Pre-Training Mandarin-L1		Post-Training English-L1		Post-Training Mandarin-L1		Reference	
Gender		M	F	M	F	M	F	M	F	M	F
Vowel	Formant	–	–	–	–	–	–	–	–	–	–
/i/	F1	390 *(19)*	438 *(37)*	376 *(4)*	423 *(49)*	363 *(22)*	405 *(24)*	359 *(35)*	410 *(54)*	384	384
	F2	1753 *(112)*	1803 *(173)*	1607 *(15)*	1900 *(97)*	1841 *(174)*	1998 *(148)*	1890 *(147)*	2090 *(218)*	2363	2409
/u/	F1	427 *(13)*	435 *(83)*	433 *(5)*	438 *(53)*	444 *(68)*	481 *(34)*	422 *(42)*	443 *(54)*	434	443
	F2	1407 *(144)*	1424 *(40)*	1219 *(105)*	1310 *(43)*	1492 *(85)*	1574 *(134)*	1124 *(22)*	1401 *(178)*	1070	1086
/a/	F1	617 *(77)*	725 *(62)*	617 *(57)*	712 *(127)*	629 *(54)*	693 *(36)*	625 *(63)*	728 *(84)*	602	708
	F2	1424 *(78)*	1669 *(70)*	1333 *(86)*	1473 *(80)*	1416 *(136)*	1598 *(168)*	1390 *(8)*	1521 *(132)*	1476	1705
/oe/	F1	459 *(22)*	557 *(28)*	461 *(21)*	493 *(58)*	461 *(35)*	548 *(25)*	446 *(1)*	545 *(47)*	509	587
	F2	1580 *(85)*	1775 *(104)*	1365 *(111)*	1601 *(38)*	1592 *(54)*	1777 *(96)*	1466 *(9)*	1661 *(85)*	1474	1676
/e/	F1	478 *(34)*	524 *(31)*	454 *(27)*	531 *(32)*	487 *(33)*	523 *(29)*	480 *(3)*	533 *(52)*	434	481
	F2	1790 *(64)*	2064 *(72)*	1668 *(90)*	1882 *(135)*	1752 *(174)*	2059 *(95)*	1788 *(133)*	2010 *(50)*	1951	2229

/ ɛ/	F1	517 *(34)*	593 *(27)*	492 *(19)*	596 *(36)*	505 *(30)*	598 *(50)*	513 *(47)*	620 *(66)*	498	571
	F2	1621 *(53)*	1831 *(128)*	1599 *(85)*	1708 *(55)*	1605 *(21)*	1854 *(99)*	1639 *(78)*	1828 *(118)*	1759	2021

The average F1 and F2 frequency values for the 6 French vowels analysed at pre and post-training for our groups separately, alongside the appropriate French reference. Mean *(SD)*.

Table 4. English Vowel Formant Frequencies

		Pre-Training		Pre-Training		Post-Training		Post-Training		Reference	
		English-L1		Mandarin-L1		English-L1		Mandarin-L1		_	_
Gender		M	F	M	F	M	F	M	F	M	F
Vowel	Formant	_	_	_	_	_	_	_	_	_	_
/i/	F1	418 *(70)*	450 *(82)*	409 *(65)*	435 *(85)*	411 *(86)*	453 *(90)*	398 *(72)*	432 *(90)*	407	444
	F2	1775 *(213)*	1934 *(280)*	1765 *(189)*	1940 *(275)*	1779 *(246)*	1933 *(393)*	1721 *(244)*	1944 *(303)*	1759	1937
/u/	F1	425 *(81)*	449 *(84)*	418 *(112)*	459 *(113)*	428 *(75)*	447 *(93)*	419 *(156)*	464 *(117)*	425	454
	F2	1500 *(234)*	1780 *(272)*	1541 *(245)*	1743 *(230)*	1533 *(274)*	1819 *(318)*	1397 *(403)*	1707 *(370)*	1488	1774
/a/	F1	615 *(73)*	725 *(87)*	600 *(65)*	715 *(84)*	620 *(85)*	732 *(98)*	596 *(58)*	704 *(108)*	612	721
	F2	1265 *(98)*	1440 *(98)*	1260 *(98)*	1395 *(90)*	1297 *(109)*	1466 *(155)*	1209 *(128)*	1343 *(101)*	1268	1417
/e/	F1	500 *(41)*	519 *(45)*	480 *(43)*	580 *(76)*	457 *(43)*	514 *(59)*	443 *(49)*	520 *(74)*	516	606
	F2	1650 *(184)*	1880 *(175)*	1700 *(162)*	1775 *(112)*	1839 *(195)*	2080 *(266)*	1765 *(124)*	2029 *(241)*	1520	1668

/ɛ/	F1	475 *(60)*	561 *(67)*	498 *(53)*	584 *(76)*	517 *(67)*	600 *(114)*	514 *(81)*	615 *(129)*	453	517
	F2	1615 *(154)*	1852 *(182)*	1623 *(159)*	1879 *(185)*	1514 *(182)*	1695 *(233)*	1524 *(166)*	1627 *(299)*	1814	2060
/ə/	F1	480 *(82)*	532 *(91)*	490 *(78)*	536 *(74)*	482 *(90)*	533 *(148)*	502 *(174)*	539 *(144)*	489	535
	F2	1567 *(221)*	1770 *(197)*	1520 *(202)*	1752 *(168)*	1580 *(254)*	1776 *(287)*	1493 *(218)*	1747 *(191)*	1551	1765

The average F1 and F2 values for the 6 English vowels analysed at pre and post-training for our groups separately, alongside the appropriate English reference. Mean *(SD)*.

IV. Voxel-Specified Neuroplasticity Analysis

Table 5. Voxel-Specified Grey Matter Volumes (mm^3) at Pre and Post Training

A. VOT Biomarker Voxels

Participant	L1	VOT	Right Cerebellum		Left dPMC		Left vPMC		Left dmPFC	
			[24, -54, -39]		[-45, 2, 39]		[-52, 2, 4]		[-9, 49, 23]	
			Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
4	E	-1	0.0089	0.0090	0.6092	0.6373	0.7736	0.7456	0.6707	0.6625
5	E	-6	0.0252	0.0172	0.4178	0.4012	0.6695	0.6644	0.6226	0.6143
6	E	-2	0.0152	0.0140	0.6519	0.6628	0.7589	0.7452	0.6785	0.6753
7	E	-2	0.0051	0.0071	0.6402	0.6026	0.7642	0.7543	0.6451	0.6348
10	E	-3	0.0118	0.0186	0.6349	0.6143	0.7404	0.7522	0.6718	0.6914
12	E	-3	0.0106	0.0064	0.6008	0.5845	0.6939	0.6908	0.6474	0.6317
14	E	1	0.0048	0.0046	0.7559	0.7474	0.8093	0.7832	0.7361	0.7379
15	E	-3	0.0044	0.0092	0.5732	0.5633	0.7627	0.7157	0.6401	0.6263
16	E	-3	0.0163	0.0173	0.709	0.7052	0.6689	0.6246	0.6806	0.6762
17	E	-4	0.0222	0.0145	0.6567	0.6679	0.7248	0.7477	0.6482	0.6605
1	M	-2	0.0175	0.0101	0.7291	0.7148	0.7842	0.7516	0.6708	0.6259
2	M	-8	0.0610	0.0448	0.3244	0.3356	0.6048	0.621	0.5197	0.5233
3	M	-3	0.0099	0.0159	0.5795	0.7089	0.6778	0.703	0.5817	0.6798
8	M	-4	0.0229	0.0414	0.5897	0.5802	0.7509	0.7656	0.6443	0.6511
9	M	-3	0.0270	0.0144	0.589	0.5731	0.7209	0.718	0.7243	0.7174
11	M	-5	0.0437	0.0420	0.4416	0.4128	0.6837	0.6699	0.6548	0.6246
13	M	-2	0.0201	0.0159	0.7349	0.7089	0.7286	0.703	0.7055	0.6798
18	M	-2	0.0110	0.0093	0.7115	0.7177	0.7231	0.7149	0.6759	0.673
Mean		-3.06	0.0188	0.0173	0.6083	0.6077	0.7245	0.7150	0.6566	0.6548

B. /i/ Change Biomarker Voxels

Participant	L1	/i/ change	Left Cerebellum		Right Cerebellum	
			[-20, -39, -32]		[19, -36, -34]	
			Time 1	Time 2	Time 1	Time 2
4	E	140.38	0.0449	0.0458	0.0666	0.0649
5	E	26.29	0.0192	0.0217	0.0280	0.0301
7	E	-17.5	0.0226	0.0267	0.0206	0.0182
10	E	201.1	0.0832	0.0713	0.1149	0.1268
12	E	130.49	0.0267	0.0327	0.0576	0.0701

14	E	155.77	0.0220	0.0195	0.0364	0.0365
15	E	94.97	0.0374	0.0435	0.0408	0.0478
16	E	543.98	0.1799	0.1402	0.1520	0.1439
17	E	28.78	0.0545	0.0265	0.0182	0.0127
3	M	187.28	0.0490	0.0808	0.0938	0.0764
8	M	376.62	0.1687	0.1454	0.1635	0.1875
9	M	-125.6	0.0422	0.0586	0.0781	0.0623
11	M	157.85	0.0750	0.0427	0.0761	0.078
13	M	29.31	0.0912	0.0808	0.0769	0.0764
18	M	209.89	0.1241	0.1186	0.1150	0.117
Mean		142.64	0.0694	0.0637	0.0759	0.0766

C. <u>Articulation Rate Biomarker Voxels</u>

Participant	L1	Articulation Rate	Left Caudate		Right IPL		dACC	
			[-16, 4, 20]		[50, -43, 38]		[-7 31 -11]	
			Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
4	E	0.33	0.0850	0.0803	0.6543	0.6489	0.6279	0.6548
5	E	-0.08	0.0238	0.0249	0.2813	0.2766	0.6283	0.6176
6	E	0.24	0.0877	0.0858	0.5370	0.5452	0.5944	0.6105
7	E	0.39	0.1063	0.1149	0.4628	0.4636	0.5422	0.4852
10	E	0.03	0.0560	0.055	0.3987	0.4198	0.7556	0.7435
12	E	0.26	0.1065	0.1065	0.5267	0.4928	0.6835	0.6809
14	E	0.04	0.0273	0.022	0.4535	0.4767	0.6473	0.6352
15	E	0.20	0.0503	0.0515	0.5709	0.5847	0.6345	0.6154
16	E	0.10	0.0526	0.0544	0.5009	0.5275	0.6354	0.625
17	E	-0.10	0.0441	0.0391	0.3166	0.3321	0.6357	0.6349
1	M	0.01	0.0431	0.0591	0.4434	0.4855	0.4765	0.4868
2	M	0.25	0.1095	0.1038	0.3188	0.3118	0.5883	0.5544
3	M	0.13	0.1106	0.114	0.2799	0.6727	0.5899	0.6535
8	M	0.30	0.0693	0.0681	0.5616	0.5784	0.5975	0.5977
9	M	0.20	0.0733	0.0706	0.5842	0.57	0.5492	0.5278
11	M	0.10	0.0811	0.0821	0.4876	0.4829	0.5332	0.5158
13	M	0.52	0.1111	0.114	0.6799	0.6727	0.6617	0.6535
18	M	0.02	0.0456	0.0408	0.4506	0.4759	0.5188	0.5137
Mean		0.16	0.0713	0.0715	0.4727	0.5009	0.6056	0.6003

D. Sentence Repetition Biomarker Voxels

Participant	L1	Sentence Repetition	Left Globus Pallidus		Left MTG (BA 21)	
			[-7, 0, 2]		[-56, -12, -17]	
			Time 1	Time 2	Time 1	Time 2
4	E	12	0.1209	0.1401	0.5369	0.5262
5	E	2	0.0683	0.0649	0.5271	0.5118
6	E	5	0.1030	0.1195	0.5223	0.5267
7	E	21	0.1486	0.1356	0.6640	0.5896
10	E	26	0.1434	0.1349	0.6857	0.6622
12	E	21	0.1363	0.1283	0.6290	0.629
14	E	3	0.0606	0.0612	0.3875	0.3982
15	E	7	0.0697	0.0824	0.5064	0.5092
16	E	11	0.0955	0.0974	0.5729	0.5514
17	E	10	0.0761	0.0752	0.5877	0.5666
1	M	-5	0.0619	0.0667	0.4208	0.4169
2	M	7	0.1057	0.0996	0.4673	0.4724
3	M	6	0.1115	0.1202	0.4795	0.4965
8	M	8	0.1007	0.0983	0.4941	0.4674
9	M	8	0.0924	0.0845	0.5468	0.5802
11	M	11	0.0753	0.0852	0.5939	0.5737
13	M	17	0.1117	0.1202	0.5039	0.4965
18	M	8	0.0930	0.0861	0.5358	0.556
Mean		9.89	0.0986	0.1000	0.5368	0.5295

E. Words per Minute Biomarker Voxels

Participant	L1	Words per Minute	Left MOG	
			[-32, -75, 3]	
			Time 1	Time 2
4	E	13.58	0.1307	0.1359
5	E	-11.63	0.0226	0.0255
6	E	5.92	0.063	0.075
7	E	31.56	0.2026	0.1957
10	E	45.09	0.2802	0.2537
12	E	21.77	0.1381	0.1372
14	E	33.85	0.2754	0.2796
15	E	14.81	0.064	0.0592
16	E	36.22	0.1193	0.1052

17	E	12.51	0.1073	0.0962
1	M	12.63	0.1119	0.1116
2	M	20.75	0.1353	0.1379
3	M	5.55	0.0483	0.1085
8	M	8.32	0.1352	0.1377
9	M	15.86	0.0968	0.096
11	M	20.60	0.1206	0.1414
13	M	12.63	0.102	0.1085
18	M	16.73	0.1469	0.1414
Mean		17.60	0.1278	0.1303

REFERENCES

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of bilinguals over monolinguals in learning a third language. *Bilingual Research Journal*, *33*(2), 173-199.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of neurolinguistics*, *20*(3), 242-275.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and cognitive processes*, *23*(4), 557-582.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., & Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral cortex*, *22*(9), 2076-2086.

Abutalebi, J., Della Rosa, P. A., Gonzaga, A. K. C., Keim, R., Costa, A., & Perani, D. (2013). The role of the left putamen in multilingual language production. *Brain and language*, *125*(3), 307-315.

Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Della Rosa, P. A., Parris, B. A., & Weekes, B. S. (2015). Bilingualism provides a neural reserve for aging populations. *Neuropsychologia*, *69*, 201-210.

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*(4), 683-695.

Arnaud, V., Sigouin, C., & Roy, J. P. (2011). Acoustic Description of Quebec French High Vowels: First Results. In *ICPhS* (pp. 244-247).

Ashburner, J., Friston, K.J., (2001). Why voxel-based morphometry should be used. *Neuroimage* 14, 1238–1243.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. Annu. Rev. Psychol., 56, 149-178.

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child development*, *83*(2), 413-422.

Barbeau, E. B., Chai, X. J., Chen, J. K., Soles, J., Berken, J., Baum, S., & Klein, D. (2017). The role of the left inferior parietal lobule in second language learning: An intensive language training fMRI study. *Neuropsychologia*, *98*, 169-176.

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive science*, *36*(6), 1129-1147.

Baum, S. R., Blumstein, S. E., Naeser, M. A., & Palumbo, C. L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*, *39*(1), 33-56.

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *Journal of Neuroscience*, *30*(45), 14964-14971.

Berken, J. A., Gracco, V. L., Chen, J. K., & Klein, D. (2016). The timing of language learning shapes brain structure associated with articulation. *Brain Structure and Function*, *221*(7), 3591-3600.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, 171-206.

Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual review of applied linguistics*, *21*, 169-181.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*(1), 27-44.

Bialystok, E., & Craik, F. I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current directions in psychological science*, *19*(1), 19-23.

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.42, retrieved 15 August 2018 from http://www.praat.org/

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2015). Language control in bilinguals: monitoring and response selection. *Cerebral Cortex*, *26*(6), 2367-2380.

Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex*, *11*(3), 230-238.

Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives*, *111*(45), 8-20.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*(1), 71-87.

Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Callahan, M., Chen, J. K., & Klein, D. (2016). Intrinsic functional connectivity in the adult brain and success in second-language learning. *Journal of Neuroscience*, *36*(3), 755-761.

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in psychology*, 5, 825.

Chen, S. A., & Desmond, J. E. (2005). Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *Neuroimage*, 24(2), 332-338.

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications*, *7*, 11070.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, 22(1), 466-476.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, *18*(2), 192-205.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, *15*(5), 336.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., & Usui, K. (2006). Language control in the bilingual brain. *Science*, *312*(5779), 1537-1540.

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D. & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. Neuroreport, 8(17), 3809-3815.

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. Science 330:1359 –1364.

DeLuca, V., Rothman, J., & Pliatsikas, C. (2018). Linguistic immersion and structural effects on the bilingual brain: a longitudinal study. *Bilingualism: Language and Cognition*, 1-16.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. Proceedings of the National Academy of Sciences, 116(15), 7565-7574.

Démonet, J. F., Price, C., Wise, R., & Frackowiak, R. S. J. (1994). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron-emission tomography study in normal human subjects. *Neuroscience letters*, *182*(1), 25-28.

Desmurget M, Turner RS (2008) Testing basal ganglia motor functions through reversible inactivations in the posterior internal globus pallidus. *J Neurophysiol* 99:1057–1076

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., & Maquet, P. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. Learning & Memory, 12(5), 480-490.

Djordjevic, J., Smith, M. L., Sziklas, V., Piper, D., Pénicaud, S., & Jones-Gotman, M. (2011). The Story Learning and Memory (SLAM) test: Equivalence of three forms and sensitivity to left temporal lobe dysfunction. *Epilepsy & Behavior, 20*(3), 518-523.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents* (No. 9). Walter de Gruyter.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA review*, *19*(1), 42-68.

Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, 59, 230-248.

Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. Neuropsychologia, 41(3), 252-262.

Eckman, F. R. (2008). Typological markedness and second language phonology. *Phonology and second language acquisition*, *36*, 95-115.

Elmer, S., Hänggi, J., & Jäncke, 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.

Feng, G., Yi, H. G., & Chandrasekaran, B. (2018). The Role of the Human Auditory Corticostriatal Network in Speech Learning. *Cerebral Cortex*.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*(1), 47-65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Floyer-Lea, A., & Matthews, P. M. (2005). Distinguishable brain activation networks for short-and long-term motor skill learning. Journal of neurophysiology, 94(1), 512-518.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914-926.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (2000). Adults' perception and production of the English vowel/i. Journal of speech, language, and hearing research, 43(1), 129-143.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. Trends in cognitive sciences, 6(2), 78-84.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, *23*(27), 9240-9245.

Gillies, M. J., Hyam, J. A., Weiss, A. R., Antoniades, C. A., Bogacz, R., Fitzgerald, J. J., ... & Green, A. L. (2017). The cognitive role of the globus pallidus interna; insights from disease states. *Experimental brain research*, 235(5), 1455-1465.

Glanz, O., Derix, J., Kaur, R., Schulze-Bonhage, A., Auer, P., Aertsen, A., & Ball, T. (2018). Real-life speech production and perception have a shared premotor-cortical substrate. *Scientific reports*, *8*(1), 8898.

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, *35*(5), 997-1010.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage*, *21*(2), 494-506.

Golestani, N., & Pallier, C. (2006). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*, *17*(4), 929-934.

Golestani, N., Molko, N., Dehaene, S., Lebihan, D., & Pallier, C. (2006). Brain structure predicts the learning of foreign speech sounds. *Cerebral cortex*, *17*(3), 575-582.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and language*, *109*(2-3), 55-67.

Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience*, 31(11), 4213-4220.

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders. Philadelphia, PA: Lea & Febiger.

Grafton, S. T., Schmitt, P., Van Horn, J., & Diedrichsen, J. (2008). Neural substrates of visuomotor learning based on improved feedback control and prediction. *Neuroimage*, *39*(3), 1383-1395.

Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral cortex*, *19*(11), 2690-2698.

Grogan, A., Jones, Ō. P., Ali, N., Crinion, J., Orabona, S., Mechias, M. & Price, C. J. (2012). Structural correlates for lexical efficiency and number of languages in non-native speakers of English. *Neuropsychologia*, *50*(7), 1347-1352.

Grosjean, F., & Li, P. (2012). The psycholinguistics of bilingualism. John Wiley & Sons.

Haspelmath, M. (2001). *Language typology and language universals: An international handbook* (Vol. 20). Walter de Gruyter.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*(3), 486-502.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135.

Hillenbrand, J. M., Houde, R. A., & Gayvert, R. T. (2006). Speech perception based on spectral peaks versus spectral shape. *The Journal of the Acoustical Society of America*, *119*(6), 4041-4054.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. Attention, Perception, & Psychophysics, 72(5), 1218-1227.

Honda, M., Deiber, M. P., Ibánez, V., Pascual-Leone, A., Zhuang, P., & Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study. *Brain: a journal of neurology*, 121(11), 2159-2173.

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. Neuropsychologia, 45(9), 1987-2008.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. Journal of Speech, Language, and Hearing Research.

Kahn, J., Audibert, N., Bonastre, J. F., & Rossato, S. (2011). Inter and Intra-speaker Variability in French: An Analysis of Oral Vowels and Its Implication for Automatic Speaker Verification. In *ICPhS* (pp. 1002-1005).

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. Nature Reviews Neuroscience, 12(4), 231.

Kantak, S. S., Mummidisetty, C. K., & Stinear, J. W. (2012). Primary motor and premotor cortex in implicit sequence learning–evidence for competition between implicit and explicit human motor memory systems. European Journal of Neuroscience, 36(5), 2710-2715.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705-710.

Kent, R. D., Read, C., & Kent, R. D. (1992). *The acoustic analysis of speech* (Vol. 58). San Diego: Singular Publishing Group.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*(4), 686-706.

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: An international journal for the rapid communication of research in Neuroscience*.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. Human brain mapping, 30(3), 859-873.

Kousaie, S., Chai, X. J., Sander, K. M., & Klein, D. (2017). Simultaneous learning of two languages from birth positively impacts intrinsic functional connectivity and cognitive control. *Brain and cognition*, *117*, 49-56.

Krashen, S. D. (1981). Bilingual education and second language acquisition theory. *Schooling and language minority students: A theoretical framework*, 51-79.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831.

Legault, J., Grant, A., Fang, S. Y., & Li, P. (2019). A longitudinal investigation of structural brain changes during second language learning. Brain and language, 197, 104661.

Lenneberg, E. 1967. Biological foundations of language. New York: Wiley and Son.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex*, *58*, 301-324.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review*, *21*(4), 861-883.

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition*, *12*(3), 251-285.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398-4403.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*.

Marian, V., & Shook, A. (2012, September). The cognitive benefits of being bilingual. In *Cerebrum: The Dana forum on brain science* (Vol. 2012). Dana Foundation.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240-244.

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation. *The Journal of the Acoustical Society of America*, *97*(3), 1865-1873.

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.

Mechelli, A., Price, C. J., Friston, K. J., & Ashburner, J. (2005). Voxel-based morphometry of the human brain: methods and applications. *Current medical imaging reviews*, *1*(2), 105-113.

Meinzer, M., Elbert, T., Wienbruch, C., Djundja, D., Barthel, G., & Rockstroh, B. (2004). Intensive language training enhances brain plasticity in chronic aphasia. *BMC biology*, *2*(1), 20.

Molaei, A. (2016). The Impact of Conscientiousness Personality Trait on Language Learning Strategy Application in EFL Context. *European Online Journal of Natural and Social Sciences*: Proceedings, 2(2s), pp-668.

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A.& Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proceedings of the National Academy of Sciences*, *95*(3), 922-929.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, 481-502.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, *66*(2), 232-258.

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., ... & Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. Science, 291(5511), 2165-2167.

Pénicaud, S., Klein, D., Zatorre, R. J., Chen, J. K., Witcher, P., Hyde, K., & Mayberry, R. I. (2013). Structural brain changes linked to delayed first language acquisition in congenitally deaf individuals. *Neuroimage*, *66*, 42-49.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

Petrides, M. (2013). *Neuroanatomy of language regions of the human brain*. Academic Press.

Pisella, L., Alahyane, N., Blangero, A., Thery, F., Blanc, S., & Pelisson, D. (2011). Right-hemispheric dominance for visual remapping in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 572-585.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*(1), 62-88.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684-1696.

Qi, Z., Han, M., Garel, K., San Chen, E., & Gabrieli, J. D. (2015). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, *33*, 14-28.

Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M., & Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: regional and individual differences. Neuroimage, 51(2), 501-511.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46-73.

Rodriguez, S. M., Archila-Suerte, P., Vaughn, K. A., Chiarello, C., & Hernandez, A. E. (2018). Anterior insular thickness predicts speech sound learning ability in bilinguals. *NeuroImage*, *165*, 278-284.

Ryalls, J., & Larouche, A. (1992). Acoustic integrity of speech production in children with moderate and severe hearing impairment. *Journal of Speech, Language, and Hearing Research*, *35*(1), 88-95.

Schlegel, A. A., Rudelson, J. J., & Tse, P. U. (2012). White matter structure changes as adults learn a second language. *Journal of cognitive neuroscience*, *24*(8), 1664-1670.

Schoonen, R., Gelderen, A. V., Glopper, K. D., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language learning*, *53*(1), 165-202.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews*, *32*(2), 265-278.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. Annual review of neuroscience, 33, 203-219.

Semel, E.H. Wiig, W.A. Secord (2003) Clinical evaluation of language fundamentals, fourth edition (CELF-4).

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of applied psychology*, *51*(3), 300.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, *17*(1), 87-97.

Sörös, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *Neuroimage*, *32*(1), 376-387.

Spinu, L. E., Hwang, J., & Lohmann, R. (2018). Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English. *International Journal of Bilingualism*, *22*(3), 350-370.

Sprung-Much, T., & Petrides, M. (2018). Morphological patterns and spatial probability maps of two defining sulci of the posterior ventrolateral frontal cortex of the human brain: the sulcus diagonalis and the anterior ascending ramus of the lateral fissure. *Brain Structure and Function*, 223(9), 4125-4152.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Lehmann, C., Wiest, R. & Dierks, T. (2009). Reduced frontal activation with increasing 2nd language proficiency. *Neuropsychologia*, *47*(13), 2712-2720.

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

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *64*(5), 1358-1368.

Swain, M., Lapkin, S., Rowen, N., & Hart, D. (1990). The role of mother tongue literacy in third language learning. *Language, culture and curriculum*, *3*(1), 65-81.

Tettamanti, M., Moro, A., Messa, C., Moresco, R. M., Rizzo, G., Carpinelli, A. & Perani, D. (2005). Basal ganglia and language: phonology modulates dopaminergic release. *Neuroreport,* 16(4), 397-401.

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead or moving past the classic model of language neurobiology. *Brain and language*, *162*, 60-71.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of cognitive neuroscience*, *18*(6), 1029-1043.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231-270.

Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953-968). Academic Press.

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289-296.

Veroude K, Norris DG, Shumskaya E, Gullberg M, Indefrey P (2010) Functional connectivity between brain regions involved in learning words of a new language. *Brain Lang* 113:21–27. CrossRef Medline

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: The Psychological Corporation.

Wise, R. J. S., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*(9158), 1057-1061.

Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain research bulletin*, *59*(2), 83-95.

Wong, P. C., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2007). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral cortex*, *18*(4), 828-836.

Wong, P. C., & Ettlinger, M. (2011). Predictors of spoken language learning. *Journal of communication disorders*, *44*(5), 564-567.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, *4*(1), 58-73.

Xiang, H., van Leeuwen, T. M., Dediu, D., Roberts, L., Norris, D. G., & Hagoort, P. (2015). L2-proficiency-dependent laterality shift in structural connectivity of brain language pathways. *Brain connectivity*, *5*(6), 349-361.

Yamamoto, K., & Sakai, K. L. (2017). Differential signatures of second language syntactic performance and age on the structural properties of the left dorsal pathway. Frontiers in psychology, 8, 829.

Yang, H., Hartanto, A., & Yang, S. (2018). Bilingualism confers advantages in task switching: Evidence from the dimensional change card sort task. *Bilingualism: Language and Cognition*, *21*(5), 1091-1109.

Yi, H. G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2014). The role of corticostriatal systems in speech category learning. *Cerebral Cortex*, 26(4), 1409-1420.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zijdenbos, A. P., Forghani, R., & Evans, A. C. (2002). Automatic" pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans. Med. Imaging*, *21*(10), 1280-1291.

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.