Age-related differences in speech and gray matter volume: the modulating role of multilingualism

NeuroImage 310 (2025) 121149

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Age-Related Differences in Speech and Gray Matter Volume: The Modulating Role of Multilingualism

Hanxiang Yu[a,b], Keyi Kang[a,b], Christos Pliatsikas[c], Yushen Zhou[d], Haoyun Zhang[a,b,*]

[a] Centre for Cognitive and Brain Sciences, University of Macau, Taipa, Macau SAR, China
[b] Department of Psychology, University of Macau, Taipa, Macau SAR, China
[c] School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK
[d] Department of English, University of Macau, Taipa, Macau SAR, China

ARTICLE INFO

Keywords:
Speech production
Aging
Multilingualism
Language entropy
Gray matter volume

ABSTRACT

Speech involves complex processes such as language formulation, motor coordination, and cognitive functions. As people age, their speech abilities often decline, showing reduced fluency and complexity. Older adults also show decreased gray matter volume. However, the relationship between age-related differences in speech and gray matter volume remain unclear. Multilinguals may exhibit unique age-related speech patterns depending on their language profiles. This study investigates the relationships between age-related differences in brain structure and multilingual speech across different languages, considering the effects of multilingual experience. An integrated measure of speech was calculated and used to reflect the overall speech quality, which was lower in older than younger adults. Native language speech (i.e., Cantonese) was better than non-native language speech (i.e., Mandarin), especially in older adults. More extensive use of multiple languages was associated with enhanced speech quality in both native and non-native languages. Age significantly impacts whole brain gray matter volume, which was lower in older than younger adults. The right middle temporal gyrus emerged as a critical region for speech in both languages in older adults. Bilateral putamen shows sensitivity to the effect of multilingual experience on speech performance in older adults. These findings underscore the complex interplay between age, multilingualism, and brain structure, providing valuable insights into the neural mechanisms underlying multilingual speech performance.

1. Introduction

Speech is a fascinating aspect of human communication, involving various processes such as language formulation, motor coordination, and cognitive functions (Dien et al., 2008; Heald & Nusbaum, 2014; Kent, 2000). As people age, it is often found that their speech ability declines, reflected by reduced fluency or complexity (Horton et al., 2010; Kemper et al., 1989; Moscoso del Prado Martín, 2017). Consistent with behavioral decline, older adults typically show structural changes in the brain, such as decreased grey matter volume (Sowell et al., 2003; Sowell et al., 2004; Winkler et al., 2003). Despite these behavioral and neural effects of aging, their direct relationship remains largely elusive. Additionally, multilinguals may exhibit distinct patterns of age-related effects on their speech in different languages depending on their multilingual profile (Gollan et al., 2010). In the current study, focusing on age-related differences, we investigated the relationships between brain structure and multilinguals' speech in different languages, considering the unique effects of multilingual experience.

Speakers vary considerably in their speech quality in both native and non-native languages, which can often be reflected in their speech fluency and complexity (Housen & Kuiken, 2009; Osborne, 2011). Fluency is a temporal feature of speech capturing rapidity and smoothness. Disfluent speech is often marked by slower and less smooth delivery (Lennon, 1990; Segalowitz & Hulstijn, 2005; Segalowttz, 2007; Yan et al., 2021). Disfluencies often include stutter-like disfluencies (e.g., elongated words, or repetition), and typical disfluencies (e.g., phonetic fragments, word/phrase revisions, or rephrase). Complexity, on the other hand, reflects the level of detail and sophistication in speech (Cheung & Kemper, 1992; Ortega, 2003). The speech complexity can be indicated by factors such as the mean length of utterances (Brown, 1973), and the mean number of clauses or verbs per utterance (Kemper et al., 1989). All these reported variables have been shown to reliably

* Address for correspondence: Haoyun Zhang, Centre for Cognitive and Brain Sciences, ICI, University of Macau, N21-1015f, Taipa, Macau SAR, China
E-mail address: haoyunzhang@um.edu.mo (H. Zhang).

https://doi.org/10.1016/j.neuroimage.2025.121149
Received 4 December 2024; Received in revised form 13 March 2025; Accepted 14 March 2025
Available online 16 March 2025

reflect speech quality (Bygate, 1999; Shriberg, 1999; Tachbelie et al., 2020).

Furthermore, the structure and quality of speech can be reflected via forming speech graphs by representing word sequences as networks, where each word is a node and the sequence is shown by directed edges (Mota et al., 2014; Mota et al., 2012). These graphs focus on recurrence patterns of words, identifying short-range (local) and long-range (global) recurrences to provide topological metrics. For instance, repeated edges (RE) is a local measure, describing the neighborhood of a node by indicating how often a word is connected to its neighbors. The largest connected component (LCC) and the largest strongly connected component (LSC) are examples of global measures, as they describe the overall connectivity and the structure of the entire network. The speech graph approach has been used to identify narrative characteristics in patients, as well as typical development population (Mota et al., 2023). In sum, there exists individual variation in speech quality and structure, which can be reflected by measures from different dimensions. Yet, most of previous studies focused on a single dimension, lacking an integrated measure to reflect overall speech quality.

From a neurocognitive perspective, as evidenced by functional and structural MRI studies, speech production often involves a broad left lateralized frontal-temporal brain network (Geranmayeh et al., 2012; Geranmayeh et al., 2014; Price, 2010), such as inferior frontal gyrus (IFG) for lexical selection (Hirshorn & Thompson-Schill, 2006; Price, 2010), middle temporal gyrus (MTG) for lexical and semantic processing (Hickok & Poeppel, 2007; Indefrey & Levelt, 2004), angular gyrus (AG) for semantic integration (Binder et al., 2009; Mirman et al., 2015; Pobric et al., 2007; Price, 2010; Visser et al., 2010), supramarginal gyrus (SMG) for phonological encoding (Indefrey & Levelt, 2000; Poldrack et al., 2001; Poldrack et al., 1999), and subcortical regions such as putamen for articulatory process (Chang et al., 2009; Price, 2010; Seghier & Price, 2009).

There also exists cross-language variation in individuals who speak more than one language (i.e., bilinguals and multilinguals, from now on, simply "multilinguals"). In single word production, multilinguals usually recruit not only language production regions but also cognitive control networks such as bilateral dorsal lateral prefrontal cortex (Abutalebi et al., 2012; Abutalebi & Green, 2007, 2008; Abutalebi et al., 2015; Korenar et al., 2023; Voits et al., 2020). Speech performance in each language is closely related to its proficiency, such that higher proficiency is associated with higher ability of speaking and understanding more fluent and complex speech (Kilman et al., 2014; Nip & Blumenfeld, 2015).

In addition to language proficiency, other aspects of multilingual experience, such as frequency and contexts of multiple languages use could also influence speech performance in each language (Gollan et al., 2011; Rosselli et al., 2000). These multifaced multilingual experiences have also been related to different brain outcomes. For instance, the Bilingual Anterior-to-Posterior and Subcortical Shift model (BAPSS, Grundy et al., 2017) posits that increasing multilingual experience leads to a shift of neural activation from frontal to posterior cortical and subcortical regions. Furthermore, the Dynamic Restructuring Model (DRM, Pliatsikas, 2020) describes neurostructural changes with increasing multilingual experience, indicating a steady volumetric increase in most subcortical structures (for example, putamen) involved in language control.

One way to systematically quantify multilingual experience is through a measure of language entropy, reflecting the diversity and uncertainty of language usage across different communicative contexts (Gullifer & Titone, 2019). Studies have reported that multilingual experience captured through language entropy significantly affects individuals' performance in language and other cognitive tasks (van den Berg et al., 2022). For instance, Kang et al. (2023) has reported faster response times during picture naming in native language related to higher language entropy, indicating that more diverse multilingual use experience was associated with a benefit in language production. Neuroimaging studies have also shown that diverse multilingual experience is associated with adaptations in brain regions such as the putamen and the cerebellum (Gullifer et al., 2021; Gullifer & Titone, 2021; Marin-Marin et al., 2022; Pliatsikas, 2020). Yet, it is not clear how the multilingual experience would modulate the relationships between brain structure and speech level productive performance in each language.

As people age, they often experience decline in language production (Burke & Shafto, 2008). Specific to speech, older adults often speak more slowly (Diaz et al., 2016; Duchin & Mysak, 1987), and produce more disfluent (Bortfeld et al., 2001; Obler & Albert, 1981), less grammatically complex (e.g., Kemper et al., 1989), and more off-topic speech (James et al., 1998). As with older monolinguals, older multilinguals also show significant age-related decline in language production, compared to younger ones. However, the deterioration speed of each language with age may differ depending on multilinguals' experience (Birdsong, 2006; Costa et al., 2012; Gollan et al., 2010; Manchon et al., 2015; Nanchen et al., 2017; Ullman, 2001).

At the brain level, older adults often display greater bilateral prefrontal brain activation during language production compared with younger adults (Diaz et al., 2016; Peramunage et al., 2011; Ralph et al., 2017; Wilson et al., 2009; Zhang et al., 2020). Older adults also show structural changes across the whole brain including the above-mentioned speech-related regions, as evidenced by reduced cortical thickness (Bilodeau-Mercure et al., 2015; Lemaitre et al., 2012), reduced white matter volume (Fotenos et al., 2005), reduced gray matter volume (Ramanoël et al., 2018) or higher gray matter atrophy (Pfefferbaum et al., 2013). Some studies have reported a direct link between gray matter volume and speech perception, with more intact gray matter structure in cognitive regions associated with enhanced speech perception (Wong et al., 2010). Older multilinguals' brains have been shown to have a layer of neuroprotection (Voits et al., 2020), and less degeneration (Lerman & Obler, 2017) compared to older monolinguals' brains. Yet, few studies have directly tested how age-related differences in brain structure would relate to speech production in different languages in multilinguals.

To summarize, a few questions remain to be addressed regarding the relationships among speech and brain structure in multilinguals' different languages in different ages. First, while acknowledging age-related differences in various dimensions of speech, there lacks an integrated measure of overall speech quality. Second, it remains largely unclear regarding the underlying brain structures related to these age differences and variations in multilinguals' different languages. To address these issues, we first calculated an integrated measure of speech, then explored age-related differences in performance in multilinguals' each language. We predicted a significant effect of age on speech performance such that older adults would show a lower score than younger adults. We also predicted to see cross-language variation in speech performance where native language would have a higher score on the integrated measure of speech. There might also be an interaction between age and language, such that the language difference would be enlarged for older adults. Moreover, more mixed use of multiple languages measured by language entropy might be associated with better speech ability in general. Furthermore, focusing on brain structures that showed age-related differences, we investigated the structural neural mechanisms related to age-related differences in speech performance and how multilingual experience modulated these relationships. We predicted that better speech performance would be associated with higher gray matter volume in the brain. Additionally, younger and older adults would show different patterns of brain-speech relationships which will also be modulated by multilingual experience captured by language entropy.

2. Methods

2.1. Participants

The study enrolled a total of 89 younger adults (18-27 years, mean = 20.9 years, SD = 1.92 years) and 31 older adults (61-76 years, mean = 67.6 years, SD = 3.84 years). All participants were native Cantonese speakers who lived in the native language environment, and additionally spoke Mandarin Chinese. Furthermore, some participants reported the knowledge with another language such as English or Portuguese. All participants self-rated proficiency for all languages they knew from four dimensions (i.e., listening, speaking, reading, and writing) on a 1-7 scale and total score ranged from 4 to 28. For younger adults, the mean proficiency was 25.05 (SD = 3.00) for Cantonese and 22.04 (SD = 4.19) for Mandarin. For older adults, the mean proficiency was 22.23 (SD =3.57) for Cantonese and 17.58 (SD = 3.99) for Mandarin. The mean proficiency of the third language was 17.25 (N = 87, SD = 3.83) for younger and 13.38 (N = 16, SD = 4.29) for older adults. There was a significant main effect of age on the proficiency of each language ($ps < .001$), and the pair-wise comparisons across all languages for each age group were all significant ($ps < .001$). All participants were right-handed and reported no neurological or psychiatric disorders. Written consent was obtained prior to the beginning of the study. All studies protocols were approved by the Research Ethics Committee of the University of Macau.

2.2. Neuropsychological Testing

Preceding the MRI session, each participant underwent a series of psychometric and neuropsychological assessments designed to evaluate their basic cognitive profiles, including processing speed, executive function, memory, and language abilities. The screening tasks included the Montreal Cognitive Assessment (MoCA, Hong Kong version) to screen out mild cognitive impairment or dementia (Nasreddine et al., 2005); a Geriatric Depression Scale with 15 items (GDS-15, a shortened version of GDS) to screen out individuals with depression (De Craen et al., 2003; Ferraro & Chelminski, 1996); a Color Vision test to screen out color blind individuals. Cognitive assessments included forward and backward digit span tasks from the Wechsler Adult Intelligence Scale (WAIS) to assess working memory (Wechsler, 1997); a simple and choice processing speed task to assess speed. Language assessment tasks included a Chinese version of a vocabulary test from WAIS Chinese version to measure vocabulary knowledge (Dai et al., 1990); a reading habits questionnaire (Acheson et al., 2008); a categorical verbal fluency task (VF) using different semantic categories as assessments of lexical retrieval in each language (Filippetti & Allegri, 2011; Friesen et al., 2015; Malek et al., 2013). The neuropsychological testing results are shown in Table 1.

In addition to the neuropsychological tasks, participants' language profile was assessed by the Language History Questionnaire Version 3 (LHQ, Li et al., 2014), capturing the linguistic background and language proficiency of multilinguals. Furthermore, all participants performed a free speech task in Cantonese and Mandarin Chinese. Specifically, the free speech task was used to elicit unrestricted verbal responses regarding language production. During the task, participants were asked to generate free speech on a certain topic. The Cantonese question was "What do you like or dislike about living in Macau", while the Mandarin question was "What do you like or dislike about summer". Participants were given 15 seconds to think about each question, then a 3-minute period to articulate their thoughts, allowing for a natural and unstructured expression of ideas. Participants' spoken responses were recorded using an audio recorder. A procedural figure including data collection, coding, and analysis can be found in Fig. 1.

Table 1
Participants' neuropsychological testing scores

	Younger adults		Older adults	
	Mean (SD)		Mean (SD)	
Cognitive Assessments				
MoCA (out of 30)	27.81 (1.85)		25.65 (2.56) ***	
Digit span forward	8.30 (2.05)		6.00 (1.73) *	
Digit span backward	5.38 (1.63)		4.35 (2.50) ***	
Simple speed (ms)	263.88 (38.94)		326.52 (87.82) ***	
Choice speed (ms)	269.77 (33.71)		405.16 (101.92) ***	
Language Profile				
WAIS Vocabulary	26.47 (5.79)		22.29 (6.30) **	
Reading habits	20.96 (4.44)		19.87 (4.46) ns	
Language Entropy	0.90 (0.32)		0.40 (0.31) ***	
	Cantonese	Mandarin	Cantonese	Mandarin
VF (correct tokens)	16.94 (7.96)	14.71 (4.14)	12.87 (3.26) ***	10.73 (2.64) ***
Speech Variables				
Speech Duration (s)	176.35 (11.83)	170.76 (21.34)	175.65 (10.97)	154.42 (34.06) *
N of Speeches< 180 s	19/89	33/89	8/31	19/31
Shortest Speech (s)	101	78	125	62
Verbs	3.89 (0.91)	2.62 (0.48)	3.69 (1.19) ns	2.08 (0.61) ***
MLU	19.70 (5.21)	17.78 (5.40)	20.10 (5.22) ns	15.45 (5.42) *
Disfluency	27.54 (9.43)	44.46 (19.56)	28.00 (5.07) ns	47.33 (13.75) ns
LCC	153.46 (35.65)	134.33 (37.19)	134.42 (27.60) **	107.71 (35.47) ***
RE	75.74 (35.37)	56.35 (34.91)	70.65 (28.63) ns	32.03 (19.11) ***
Speech Factor	.57 (.85)	-.29 (.70)	.25 (.80) ns	-1.06 (.77) ***

Reported are means with standard deviations in parentheses, unless otherwise noted. Age difference was marked as: ***, $p < .001$; **, $p < .01$, *, $p < .05$, ns indicates not significant. Abbreviations, VF, Verbal Fluency; Verbs, Verbs per Utterance; MLU, Mean Length of Utterances; LCC, Largest Connected Component; RE, Repeated Edges. Variables in bold were included in calculating the speech factor.

2.3. MRI Data Acquisition

Following the completion of the neuropsychological testing, participants were invited to take part in the MRI scan. MRI data were collected on a 3T Siemens Prisma MRI scanner with a 32-channel head coil. We collected a sagittal T1 weighted localizer image to define a volume for data collection and higher-order shimming. The anterior and posterior commissures were identified for slice selection and shimming. T1 weighted structural images were then collected using a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (repetition time [TR] = 2300 ms; echo time [TE] = 2.28 ms; Inversion Time [TI] = 900 ms; flip angle = 8°; echo spacing = 7 ms; acceleration factor = 2; field of view [FOV] = 256 mm^2; voxel size = 1 × 1 × 1 mm; 160 contiguous slices).

2.4. Data Coding

2.4.1. Language entropy

Language entropy is a method for accessing the language usage diversity for multilinguals (Gullifer et al., 2018; Gullifer et al., 2021;

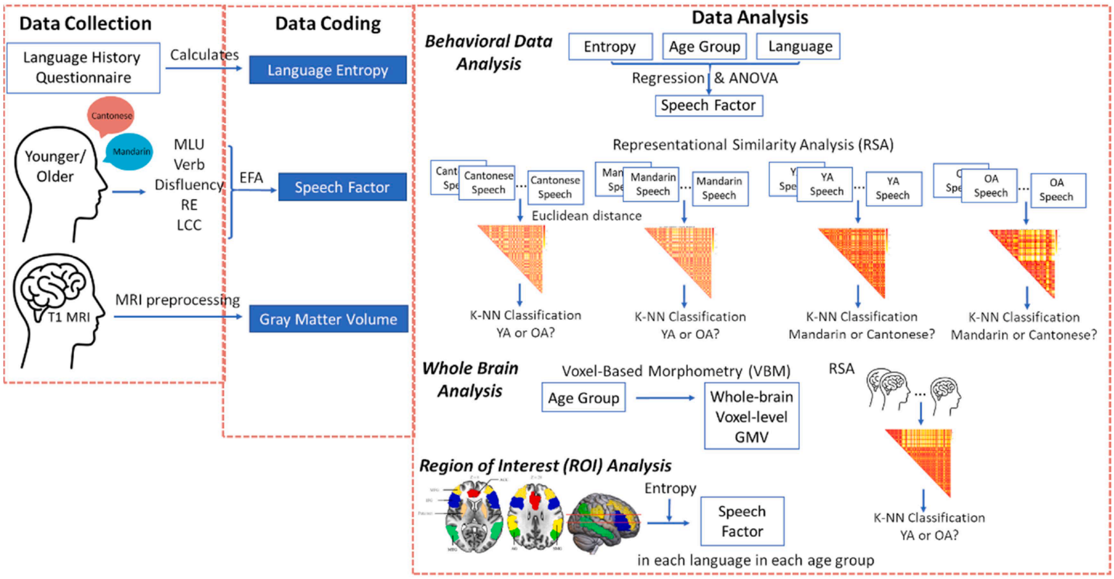

Fig. 1. Procedural figure summarizing data collection, coding and analysis steps.

Gullifer & Titone, 2019, 2021). Language entropy was developed based on the Shannon entropy, using the equation: $H = -\sum_{i=1}^{n} P_i log_2(P_i)$ (Shannon, 1948). In this equation, n means the number of languages used in that context and P_i represents the proportion that each language$_i$ is used within a context. In the present study, language entropy was calculated based on three questions from the LHQ assessing the language use in 17 contexts (Li et al., 2014). In these questions, participants were asked to evaluate the frequency of using each language in self-engaged activities (e.g., self-talk, pray, remember numbers) on a 7-point scale; the time spent in each language per external activity (e.g., watching television, using social media, writing, reading); or when speaking with different people (e.g., family member, colleague, classmate). The calculation was conducted with the R software using the *LanguageEntropy* package (Gullifer & Titone, 2018). The language entropy of each context was calculated first, and the mean entropy across all contexts was calculated, with each participant had one language entropy score.

2.4.2. *Disfluency and complexity in free speech*

Speech data in Cantonese and Mandarin was first transcribed using the Computerized Language Analysis (CLAN, version 25) software (Macwhinney, 2000). Transcription was conducted by several research assistants and one of them inspected the coding for all participants, to make sure that the coding criteria were consistent. Although all participants were encouraged to speak for the whole 3 minutes, some were not able to do so, therefore, the duration of speech varied across both age groups and languages, as reported in Table 1.

While transcribing, the following events were coded respectively with certain symbols, including irregular pauses (unexpected breaks within the utterance; e.g., I like (.) psychology); filled pauses (use of um, er, uh, etc.; e.g., I like &-um psychology); phonetic fragments (partial words; e.g., I like &+psycho psychology); elongated words (stretching out sounds; e.g., I like: psychology); simple repetition (repeating a word or phrase; e.g., I <like> [/] like psychology); rephrasing (starting an utterance, then rephrasing part/all of it; e.g., < I like > [//] I like psychology); unintelligible speech (e.g., I xxx psychology) and the errors that were not self-identified and corrected (e.g., I liked [:like] [*] psychology). After transcription, morphological coding was conducted using the *MOR* function (Macwhinney, 2000) and lexicons provided by *TalkBank* (Macwhinney, 2000). If a word produced by participants was not included in the lexicon but indeed a real word, it was then added to the lexicon of the corresponding language.

The speech quality was first evaluated based on three variables calculated from CLAN, namely the Mean Length of Utterances (MLU), number of verbs per utterance (Verbs), and percentage of disfluencies (Disfluency; Macwhinney, 2000). Specifically, MLU is a measure used to assess speech complexity, calculated by dividing the total number of morphemes by the total number of utterances. To further quantify the complexity of sentence structure, the mean number of verbs per utterance was calculated, by including regular verbs, copulas (e.g., run, know, be), and past or present participles, but not including modal words (e.g., will, might). Furthermore, the percentage of disfluencies was calculated as the sum of stutter-like and typical disfluencies divided by the total number of syllables.

2.4.3. *Graph-based measures in free speech*

In addition to using traditional measures to evaluate speech quality, we innovatively visualize and quantify speech data with a graph approach, using *SpeechGraph* software (Mota et al., 2014; Mota et al., 2012). Specifically, *SpeechGraph* uses the transcribed speech text as input and produces a graph, with words being nodes and connected words forming pathways. The largest connected component (LCC) refers to the total number of nodes in the largest sub-graph where each node is connected to every other node through a path. The LCC has been found to be associated with aging (Malcorra et al., 2021). Additionally, repeated edges (RE) is calculated by summing all edges linking the same pair of nodes. Higher RE is associated with longer utterances, but could also indicate lower lexical diversity in speech when controlling for the length of utterances.

2.4.4. *MRI data preprocessing*

The preprocessing of T1 structural MRI images was carried out using Statistical Parametric Mapping 12 (SPM12; Wellcome Department of Cognitive Neurology, University of London) and the CAT12 toolbox (Computational Anatomy Toolbox; C. Gaser, Jena University Hospital, Jena, Germany), operated on MATLAB (The MathWorks, Natick, MA). Specifically, the structural images were segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid, and then transferred into MNI coordinate system (Montreal Neurological Institute). Subsequently, the images were normalized to a standard anatomical template using Dartel registration (Ashburner, 2007) and modulated in a non-linear way by using Jacobian determinants. Finally, a Gaussian kernel was applied to smooth the images, with an 8 mm full-width at half maximum (FWHM) in SPM12 to improve the signal-to-noise ratio and accommodate anatomical variability. The final output images were in resolution of 1.5 × 1.5 × 1.5 mm voxel size.

2.5. Data Analysis

2.5.1. Factor analysis on speech performance

As mentioned earlier, all participants performed a free speech task in Cantonese and Mandarin and their performance was evaluated by five variables, namely MLU, Verbs, Disfluency, LCC, and RE. Although the speech length varies across languages and age groups, these variables were derived from averaged features across all utterances, ensuring that speech length did not impact the assessment of speech characteristics. To reflect the overall speech quality, an exploratory factor analysis was conducted across all speech variables in both languages across all participants. There was no missing data and outlier identified based on Mahalanobis Distances (Probability $< .001$). There was no multicollinearity concern among the cognitive variables as assessed by Variance Inflation Factor (VIFs, < 3.5) and the data were normally distributed. A Bartlett's test was conducted to determine the correlation adequacy among variables from those cognitive tasks, and a Kaiser-Meyer-Olkin test (KMO, Kaiser, 1974) was then conducted to determine the sampling adequacy. Results suggested that there was a substantial correlation among the cognitive variables (Bartlett test $p < .001$) and the sample was adequate (KMO $= .79 > .60$, Kaiser, 1974), which motivated the factor analysis. All speech variables were standardized using the *scale()* function in the R environment ((score-mean)/sd). One factor was extracted based on the *kaiser* criterion. The final model used the *varimax* rotation and accounted for 58% of the variance in the data (TLI: 0.80; CFI: 0.90; RMSR: .07; RMSEA: 0.23). The *psych* package in the R environment was used for the factor analysis (Revelle, 2015).

After the latent factor was identified, the standardized speech factor score was then calculated for each language for each participant, with an overall mean across both languages and all participants to be 0. A positive speech factor score indicates above-average speech ability, while a negative score suggests below-average speech ability. The speech factor scores between the two languages were highly correlated, in both younger adults ($r = .69$, $p < .001$) and older adults ($r = .78$, $p < .001$). On speech factor scores, multi-level regression analyses were first conducted to explore the effects of Language (Cantonese vs. Mandarin), Age Group (Younger vs. Older), and multilingual experience (reflected by language entropy) on speech performance. Categorical variables were contrast coded with -1 vs 1. In the cases where interactions were significant, further analyses were conducted to clarify dynamic interplay among different variables on speech factor. Furthermore, because the sample size was unbalanced in younger and older adults, permutation tests were further conducted. Specifically, when needed, the same test was repeated for 1000 times with a random subset of younger adults each time, equivalent to the sample size of older adults.

In addition to the traditional method, the differences between age groups and between languages were explored using a classification approach based on similarity matrices generated from Representational Similarity Analysis (RSA, Kriegeskorte et al., 2008). To investigate age differences, a similarity matrix was created for each language using the speech factor scores. The matrix was calculated from the pairwise Euclidean distances across all participants. Longer distance would indicate more different speech scores between every pair of participants. For each language matrix, a classification was conducted using the k-Nearest Neighbors (k-NN) algorithm. Each similarity matrix was split into training (~70%) and testing (~30%) sets, ensuring the same proportion for each group. A k-NN classification was performed with $k = 3$. The classification performance was evaluated using a confusion matrix and accuracy calculation, indicating how accurately individuals can be classified into their respective age groups. Similarly, to explore language differences, a similarity matrix was generated for each age group separately based on the speech factor scores, using pairwise Euclidean distance across both languages. A similar classification test was then conducted for each group to determine how accurately data points could be classified in to the corresponding language.

2.5.2. Whole brain analysis

To explore the age difference on the whole brain level, we first conducted a voxel-based morphometry (VBM) analysis, focusing on the gray matter volume using SPM12. With the VBM approach, we explored the main effect of age group on whole brain GMV, including the Total Intracranial Volume (TIV) as a covariate to accommodate the variation of participant's head size. Clusters were considered significant at $p < .05$ after correcting for multiple comparisons using the Family Wise Error Rate (FWE).

In addition to the VBM approach, an RSA-based classification test was conducted on GMV, exploring potential age differences across the whole brain. Specifically, the GMV was extracted from 160 Regions of Interest (ROIs) based on the Automated Anatomical Labeling atlas 3 (AAL3; Rolls et al., 2020). Then the pairwise Euclidean distances across all participants for all ROIs were computed to form a distance matrix, which was then converted into a similarity matrix. A classification test, similar to the behavioral analyses, was performed on this GMV similarity matrix. All these analyses were conducted in the R environment (R Core Team, 2013).

2.5.3. Region of interest analysis

In addition to the whole brain analysis, several critical regions of interest (ROI) were identified to further explore the contribution of these brain structures to speech performance and the modulation of multilingual experience. Focusing on regions that showed lower GMV in older adults compared to younger adults, and combined with previous literature documenting regions important for language and cognitive functions, a total of 14 bilateral ROIs related to language and cognitive functions were selected (Fig. 2). These ROIs were left and right anterior cingulate cortex (ACC), middle frontal gyri (MFG), inferior frontal gyri (IFG), supramarginal gyri (SMG), angular gyri (AG), middle temporal gyrus (MTG), and putamen. The GMV of all ROIs were extracted using the AAL3 atlas.

With the GMV extracted from these ROIs, we analyzed their main effects and interaction with language entropy on speech factor score in each language in each age group, using the Generalized Additive Mixed Models (GAMMs) to accommodate both the linear and non-linear relationships while controlling for the effect of TIV. The *gam* function from the *mgcv* package (Wood, 2011) in the R environment was used to fit all GAMM models. For all the interested variables, the *p*-values of model fit were higher than .05, and k-indexes were close to 1, suggesting that there were no significant or missed patterns in the residuals of our models.

3. Results

3.1. Behavioral analysis on speech factor

An exploratory factor analysis identified one speech factor across all individual language variables. Specifically, the identified speech factor loaded positively on Verbs (loading $= .80$), MLU (loading $= .91$), LCC (loading $= .75$), RE (loading $= .73$), and negatively on Disfluency (loading $= -.59$). Therefore, higher speech factor score would indicate better speech quality in general.

We first explored the effects of Language, Age Group, and multilingual experience (reflected by language entropy) on speech performance (Fig. 3A), as measured by the speech factor score. Starting with the full multi-level regression model, which included all main effects and interaction combinations, we employed a stepwise approach to compare the goodness of fit of various statistical models by sequentially dropping variables. The model with lowest Akaike Information Criterion (AIC, indicating a good fit with fewer parameters) was kept. The final model included the main effects of Language, Age Group, Entropy, and the interaction between Language and Age Group. The main effect of

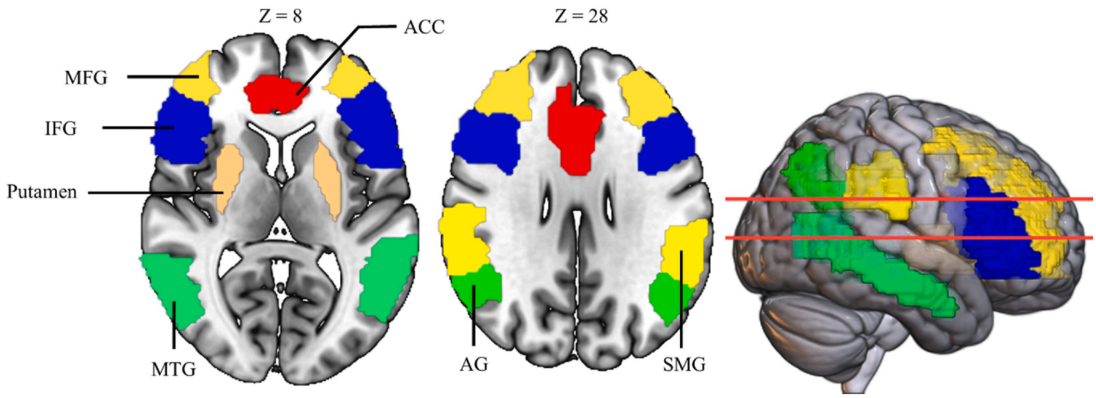

Fig. 2. Selected 14 bilateral ROIs based on the AAL3 atlas.

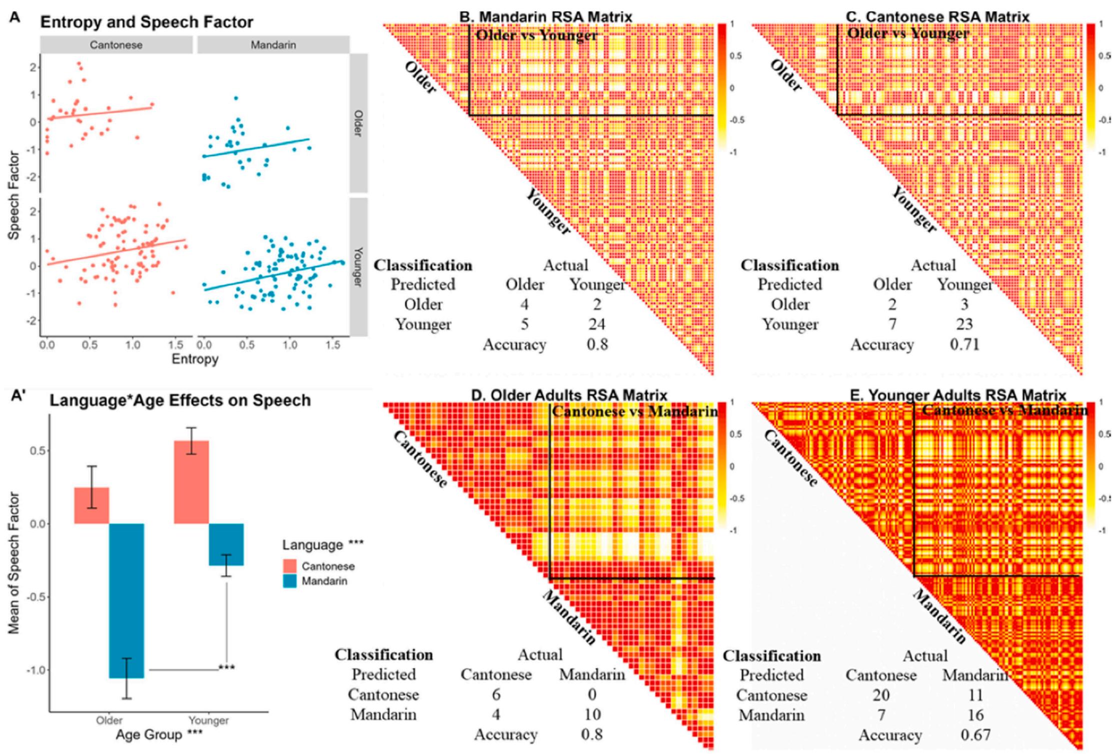

Fig. 3. Behavioral results. A) Regression analyses indicated main effects of entropy, language, and the interaction between language and age group. A') The interaction between language and age group was further explored with ANOVA, showing significant main effects of language and age group, and their interaction (*** indicates $p < .001$). The right panels show RSA matrices created from the pairwise similarity comparison for all participants in Mandarin (B) and Cantonese (C), and for both languages in Older (D) and Younger (E) adults. Data from the same age/language group (upper left and bottom right sections) showed higher level of similarity on each matrix compared to data from different age/language groups (upper right section). The bottom left panel shows the confusion matrix for this RSA matrix classification performance.

Language was significant, such that the speech factor score in Cantonese was higher than Mandarin, $\beta = .54$, $t = 17.27$, $p < .001$. While the main effect of Age Group was not significant in the regression ($\beta = .12$, $t = 1.35$, $p = .18$), its interaction with Language was significant ($\beta = .11$, $t = 3.64$, $p = .0004$). Additionally, the main effect of entropy on speech factor score was significant, such that higher entropy was significantly associated with better speech performance, $\beta = .23$, $t = 3.09$, $p = .003$. Because there were unequal number of participants in each age group, we further conducted a permutation test, randomly selecting a subset of younger adults each time to match with the sample size of older adults, repeated for 1000 times. The permutation test indicates stable significant effects of Language ($p < .001$, with fewer than 1 in 1000 of the shuffled datasets produced non-significant effect of Language), and its interaction with Age Group ($p = .046$, with only 46 in 1000 of the shuffled datasets produced non-significant interaction effect). Yet, the effect of entropy was not consistently shown in permutations ($p = .54$, implying that the observed effect could potentially be related to random variation in the data).

To further clarify the interaction between Language and Age Group on speech performance, a mixed-ANOVA combined with post-hoc pairwise comparisons was conducted, focusing on the effects of language and age group on speech factor score (Fig. 3A'). Consistent with the regression analysis, the main effect of Language was significant, such that Cantonese speech factor score was higher than Mandarin, $F(1, 118) = 298.30$, $p < .001$. Interestingly, the main effect of Age Group was also significant, such that the factor score of younger adults was higher than older adults, $F(1, 118) = 13.12$, $p < .001$, although the age effect was not significant in the initial regression. Additionally, the interaction between Language and Age Group was still significant, $F(1, 118) = 13.22$, $p < .001$. To explore the interaction, post-hoc analyses were conducted using the Tukey method to adjust for multiple comparisons. Significant higher speech factor scores were found in Cantonese than Mandarin in younger adults ($p < .001$), as well as in older adults ($p < .001$). Additionally, younger adults showed higher speech factor scores than older adults in Mandarin ($p < .001$), but not in Cantonese ($p = .27$). Given the unequal sample size, similar permutation tests were conducted for the ANOVA analysis. The permutation test indicates stable significant effects of Language ($p < .001$, suggesting that this effect is highly unlikely to have occurred by chance), Age Group ($p = .046$, with only 46 in 1000 of the shuffled datasets produced non-significant effect of Age Group), and their interaction ($p = .046$, with only 46 in 1000 of the shuffled datasets produced non-significant interaction effect), consistent with the earlier reported ANOVA results.

Last but not least on the speech performance, we used a k-NN classification test to further explore the interaction effect of Age Group and Language, based on RSA matrices. For each matrix, the training set included 70% data and the test set included the remaining 30% data. In the test set of Mandarin speech exploring age effects (9 older and 26 younger participants), 4 older adults and 24 younger adults were correctly classified. There were 5 misclassifications for older adults and 2 for younger adults. The overall classification accuracy was 80%, demonstrating a good level of accuracy in predicting age group membership based on Mandarin speech (Fig. 3B, upper left and bottom right for same age groups, upper right for different age groups). On the other hand, in the test set of Cantonese speech exploring age effects (9 older and 26 younger participants), 2 older and 23 younger adults were correctly classified. However, 7 older adults were misclassified, and 3 younger adults were misclassified. The overall classification accuracy was 71.43%, demonstrating moderate accuracy in predicting age group membership based on Cantonese speech (Fig. 3C). These results indicate that the model correctly classified the majority of younger adults but struggled with correctly classifying older adults, suggesting a less clear boundary between age groups on Cantonese speech, consistent with the regression and ANOVA results.

Additionally, focusing on language difference, classification tests were conducted in each age group separately. In the test set of older adults (10 data points for each language), 6 Cantonese speech and 10 Mandarin speech were correctly classified. There were 4 misclassifications for Cantonese. The overall classification accuracy was 80%, demonstrating a good level of accuracy in predicting language type based on older adults' speech (Fig. 3D, upper left and bottom right for same language, upper right for different languages). Finally, in the test set of younger adults (27 data points for each language), 20 Cantonese speech and 16 Mandarin speech were correctly classified. Yet, there were 7 misclassifications for Cantonese and 11 for Mandarin. The classification accuracy was 66.67%, indicating a low accuracy in predicting language membership based on younger adults' speech (Fig. 3E). These results suggest that older adults' Cantonese and Mandarin speech were more different while younger adults' proficiency in two languages were more similar, consistent with significant interaction between Language and Age Group from the multi-level regression and ANOVA analyses.

3.2. Whole brain analysis results

Whole brain VBM analyses were first conducted to investigate the group difference on GMV (see Fig. 4A and Table 2). Compared with younger adults, older adults showed lower GMV throughout the majority regions of the brain, and higher GMV in occipital lobe, bilateral thalamus and pallidum.

Additionally, with the GMV RSA matrix created from 160 ROIs across the whole brain for each participant, a k-NN classification was conducted to further explore age difference. The training set included 70% data (22 older and 63 younger participants) and the test set included the remaining 30% data (9 older and 26 younger participants). In the test set of GMV data, 7 older adults and 23 younger adults were correctly classified. There were 2 misclassifications of older and 3 misclassifications of younger adults. The overall classification accuracy reached 85.71%, indicating that the GMV similarity patterns (Fig. 4B) are effective in distinguishing different age groups.

3.3. ROI analysis results

Based on the critical regions that showed significant age effects, we further explored the main effects of the GMV in each ROI and their interactions with language entropy on speech factor scores in different languages of each age group, controlling for individual differences in head size via TIV (Fig. 5). In older adults, there was a significant nonlinear main effect of the left Putamen on Cantonese speech (Fig. 5A; smooth term edf = 3.65, ref.df = 3.92, F = 8.13, $p = .0005$), indicating a fluctuating relationship (starting from negative, then positive, then negative) between GMV and speech performance. There was also a significant interaction between the left Putamen and entropy on older adults' Cantonese speech (Fig. 5A; smooth term edf = 1.00, ref.df =

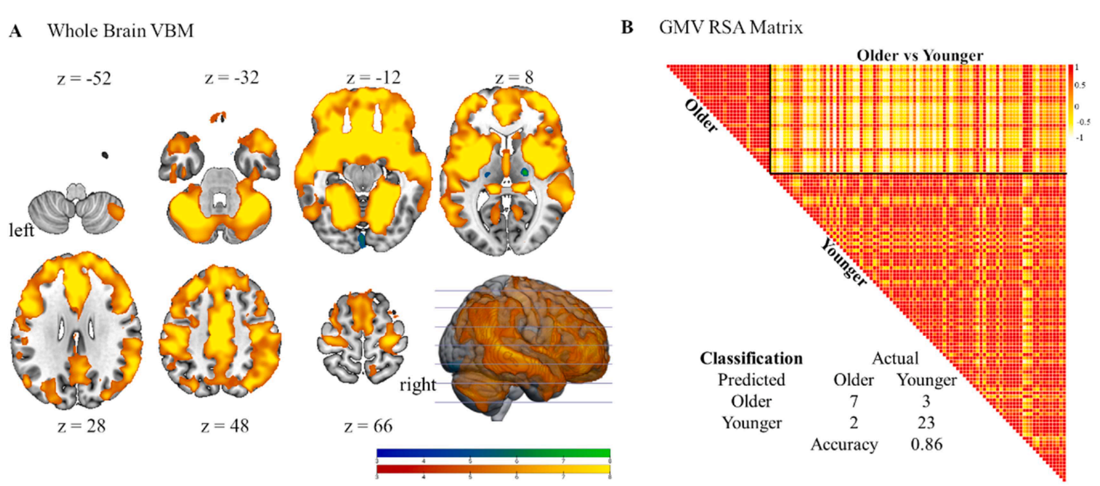

Fig. 4. Main effect of age group on whole brain GMV. A) Age comparison with VBM approach. Regions in red-yellow indicate higher GMV in younger than older adults. Regions in green-blue indicate higher GMV in older than younger adults. Color bar indicates t values. B) RSA Matrix created from the pairwise similarity comparison on GMV across whole brain for all participants. Participants from the same age group (upper left and bottom right sections) showed higher level of similarity on each matrix compared to participants from different age groups (upper right section). Classification confusion matrix shows accuracy of 86%.

Table 2
Main effect of age group on whole brain GMV with VBM approach.

	Hemisphere	Voxel	MNI coordinates (mm)			T value
			x	y	z	
Younger > Older						
Putamen	Right	261684	34	0	3	14.89
	Left		-33	-4	-9	14.84
MFG	Right		37	55	8	8.84
	Left		-37	55	8	8.51
IFG	Right		48	42	8	9.22
	Left		-48	42	8	7.39
ACC	Right		8	42	23	8.24
	Left		-8	42	23	9.41
SMG	Right		57	-41	33	5.82
	Left		-58	-26	33	4.96
MTG	Right		63	-32	0	8.41
	Left		-63	-32	0	7.14
AG	Right		62	-59	29	6.79
	Left		-55	-66	29	5.71
Occipital cortex	Left	17	-27	-86	10	5.30
Older > Younger						
Thalamus	Right	210	20	-18	3	8.22
Occipital pole	Middle	255	0	-93	-16	7.96
Thalamus	Left	110	-20	-20	3	6.96
Parahippocampal gyrus	Left	37	-16	-2	-39	5.83

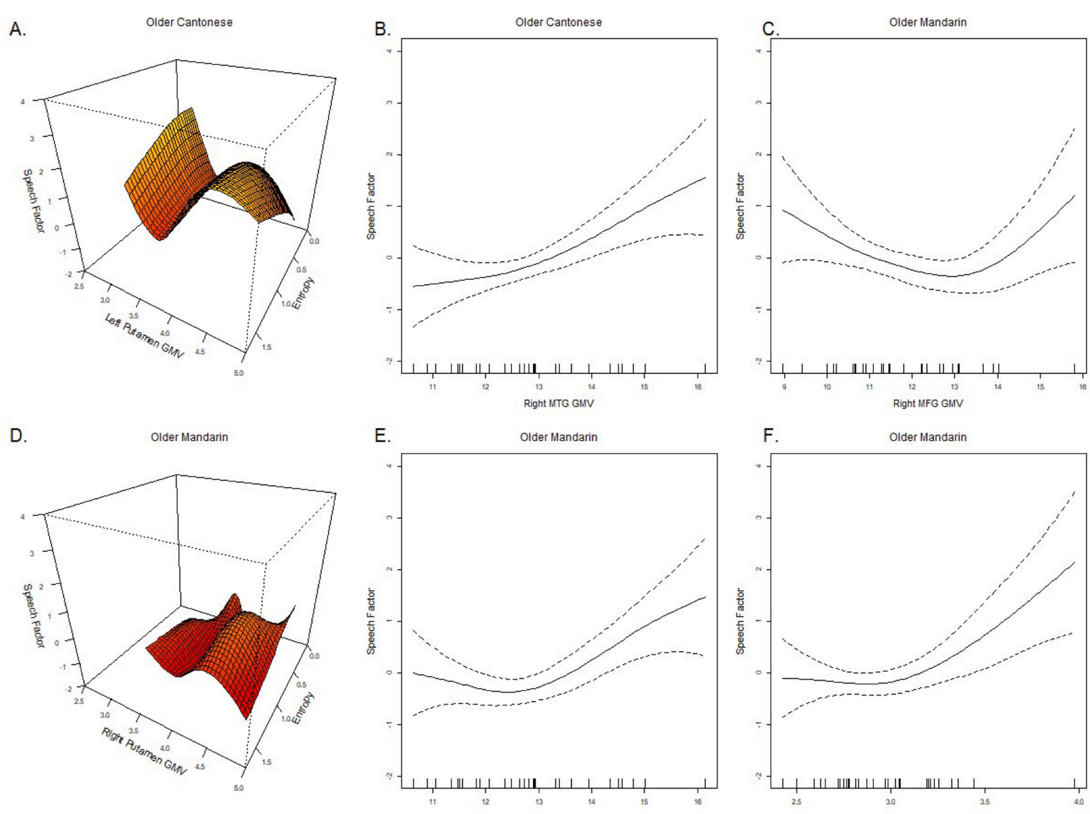

Fig. 5. Effects of ROI GMV and language entropy on Speech factor in older adults. A) shows significant main effect of the Left Putamen GMV and its interaction with entropy on Cantonese speech. B) shows the main effect of the Right MTG GMV on Cantonese speech. C) shows the main effect of the Right MFG GMV on Mandarin. D) shows the interaction between Right Putamen GMV and entropy on Mandarin speech. E) shows the main effect of the Right MTG GMV on Mandarin speech. F) shows the main effect of Left SMG GMV on Mandarin speech.

1.00, F = 4.67, $p = .04$). Specifically, the relationship between the GMV of the left putamen and the speech factor follows a horizontal "S" shape, and this shape gets steeper when entropy was lower. Additionally, participants with lower GMV showed a negative relationship between entropy and speech performance while those with higher GMV showed a positive relationship between entropy and speech performance.

In addition to the left putamen, there was a significant interaction effect between the right Putamen and entropy on older adults' Mandarin speech (Fig. 5D; smooth term edf = 5.62, ref.df = 6.88, F = 3.40, $p = .02$). Specifically, the relationship between the right Putamen and Mandarin shows a "S" shape for individuals with higher entropy. Yet, for individuals with lower entropy, the relationship becomes more complicated with multiple waves of fluctuations. Additionally, individuals with higher right Putamen GMV showed a stronger positive correlation between language entropy and Mandarin speech.

Furthermore, there was a significant positive exponential effect of the right MTG on older adults' Cantonese speech (Fig. 5B, smooth term edf = 1.65, ref.df = 1.94, F = 6.79, $p = .005$), such that higher GMV was associated with better speech performance and this relationship is accelerated when GMV is higher. For older adults' Mandarin, there was a significant non-linear exponential effect of the right MTG on speech (Fig. 5E; smooth term edf = 2.28, ref.df = 2.75, F = 3.49, $p = .03$), similar to the patterns identified in Cantonese.

Moreover, in older adults, a significant non-linear effect of the right MFG on Mandarin speech was observed (Fig. 5C; smooth term edf = 2.56, ref.df = 3.04, F = 3.48, $p = .03$), indicating a U shape (starting from negative then become positive). There was a significant non-linear effect of the left SMG on older adults' Mandarin speech (Fig. 5F; smooth term edf = 2.33, ref.df = 2.79, F = 4.18, $p = .02$), such that higher SMG was associated with better speech, and this relationship is stronger with higher GMV. It is important to note that, however, none of the effects above survived correction for multiple comparisons. There was no significant main effect of entropy identified ($ps > .1$). And there was no significant relationship between GMV of critical ROIs and speech performance in either language in younger adults ($ps > .1$).

4. Discussion

The present study investigates the relationships between speech and brain structure with a focus on age-related differences and multilingualism. Brain structure was quantified as gray matter volume. A comprehensive speech factor was calculated through a factor analysis, reflecting overall speech quality. In addition to the cross-language comparison, multilingual experience was captured through language entropy. Our results showed that higher entropy was associated with better speech performance. There was also a moderate superior performance in younger than older adults, a stable superior speech in native than non-native language, especially in older adults. Age significantly affected the whole brain gray matter volume such that older adults showed lower GMV than younger adults. Furthermore, the right MTG emerged as a critical region for speech in older adults. Additionally, subcortical regions such as the putamen were sensitive to the effect of multilingual experience on speech performance in older adults. Below, we discuss these findings in details.

Behaviorally, with the integrated speech factor reflecting both fluency and complexity (Housen & Kuiken, 2009; Osborne, 2011), we found that older adults tend to produce speech with lower quality across both languages than younger adults. Previous studies have often reported word retrieval failures with age, in not only the word level (Burke & Shafto, 2011; Diaz et al., 2016), but also the context level (Bortfeld et al., 2001; Mortensen et al., 2006). Specific to the context level, older adults tend to produce less fluent and complex speech, consistent with the current study. However, the effect of age on speech was not significant in the initial regression which added the effect of entropy, indicating a potential compensatory effect from multilingual experience to age effects on speech performance. In fact, there was a significant difference in language entropy between the two age groups, with older adults exhibiting lower entropy than younger adults ($p < .001$; Fig. 3A). This suggests that the effect of age group may be partially confounded by the effect of language entropy, which we acknowledge as a potential limitation of the current study (discussed in later sections).

Focusing on the interplay between multilingualism and age groups, there was a significant effect of language type, such that Cantonese speech quality was higher than Mandarin, which is expected given Cantonese was the native and dominant language of the current sample. There was also a main effect of language entropy, across all participants' both languages. Specifically, higher language entropy was significantly associated with higher speech factor scores, indicating that greater multilingual diversity may enhance overall speech performance for both native and non-native languages (Fig. 3A). These results suggested that the practice of frequently using multiple languages might be beneficial for language production ability, consistent with prior research (Kang et al., 2023).

Importantly, a stably significant interaction between language and age group on speech performance was also found. Specifically, the cross-language difference in speech was more pronounced in older adults compared to younger adults. This finding was further supported by the RSA-based classification test, which showed higher classification accuracy for differentiating the two languages in older adults than in younger adults. These results suggest that older adults' proficiency in the two languages was more distinct, while younger adults' proficiency was more similar. Additionally, younger adults had higher speech factor scores in Mandarin compared to older adults, but no significant age differences were observed in Cantonese. This stronger age difference in Mandarin was also further supported by the RSA-based classification test, which accurately differentiated between younger and older adults' Mandarin speech, but not Cantonese. These effects suggest that age-related differences in speech performance may be more pronounced in the non-native language. Yet, this result only speaks to the current sample because older adults' Mandarin proficiency was lowest among others. These samples may not fully represent the broader population of multilingual older adults with more balanced language proficiency. Further studies should recruit more balanced multilinguals in both age groups to further test the pronounced age difference in non-native languages.

Neurally, focusing first on age-related differences in whole brain gray matter volume, we found that older adults exhibited reduced GMV compared to younger adults throughout the entire brain. This finding was supported by both a traditional VBM approach and a classification test based on the GMV representative similarity matrix, which highly accurately differentiated between younger and older adults. The age-related brain atrophy has been reported in many studies, especially in anterior regions (Hafkemeijer et al., 2014; Raz et al., 2005; Resnick et al., 2003), associating with worse cognition (Zimmerman et al., 2006). Yet, few studies have investigated the contribution of reduced gray matter volume to age-related difference in speech production. In this context, the current study explored the age-related relationships between GMV and speech performance, as well as the modulation of multilingual experience. Several key results should be highlighted.

First, from regions that showed significant age differences in GMV, we identified several ROIs critically involved in language processing and executive functions (Papeo et al., 2019; Pliatsikas, 2020; Turker et al., 2023). Among these regions, only the right MTG, right MFG, left putamen, and left SMG showed significant contributions to speech factors in older adults, highlighting the importance of these regions to older adults' speech. The left putamen, which plays a key role in speech sequencing, and the left SMG, involved in processing sublexical information, have both been shown to be critically engaged in language processing (Tremblay & Deschamps, 2016). Unlike the left-lateralized pattern often observed in younger adults, the involvement of bilateral regions suggests a shift towards more bilateral processing, compensating for declines in the left hemisphere's language regions (Cabeza, 2002). For example, while the left MTG is generally associated with semantic processing (Indefrey & Levelt, 2000), the activation of the right MTG in older adults could reflect compensatory mechanisms for linguistic tasks. Additionally, the right MFG, known for its role in executive functions and cognitive control (Ridderinkhof et al., 2004), may become more engaged in older adults due to the increased need for cognitive resources when processing non-native languages. However, it is important to note that none of these effects survives multiple comparison correction, meaning that these findings should be interpreted with caution.

Furthermore, for older adults' both languages, the indicator of multilingual experience (indicated by entropy) dynamically modulated the relationships between GMV in putamen and speech factor in older adults. The critical role of putamen in language processing has been shown in previous studies (Turker et al., 2023; Viñas-Guasch & Wu, 2017), and in older adults (Tremblay & Deschamps, 2016). Interestingly, the divergent lateralization patterns observed in our study, with the left portion involved in native language processing and the right portion involved in non-native language processing, offer valuable insights into the neural mechanisms underlying multilingual speech. A previous connectivity study reported that the left putamen coactivated mainly with left hemisphere clusters directly associated with language processing, while the right putamen coactivated with regions involved in broader semantic and memory processing (Viñas-Guasch & Wu, 2017). Informed by this study, the pattern identified in the current research suggests that older adults' native language processing relies more heavily on left hemisphere language-specific regions, whereas non-native language processing appears to engage right hemisphere domain-specific regions. However, it is important to note that since only the putamen, but not other critical language-related regions, exhibited divergent lateralization patterns for native and non-native languages, these results should be interpreted with caution.

Yet, no significant effect was found on the relationships between speech factor and GMV in critical ROIs in younger adults. This result may suggest that proficient non-native language processing in younger adults is largely supported by their native language system. Younger adults' brains may already process both languages effectively, so the structural characteristics required for non-native language processing may not be as distinct or dependent on additional neural resources. An alternative explanation is that, although younger adults may score better on proficiency measures, their bilingual experience may not have been extensive enough to trigger the brain expansion and normalization process that typically occurs over time in older multilinguals. These combined results suggest that specific multilingual experiences play a crucial role in modulating neural bases, ultimately affecting speech production, especially in the aging brain (Pliatsikas, 2020).

While our findings are novel and informative, several limitations should be acknowledged. First, the unequal number of participants across age groups may affect the generalizability of our findings, as well as the comparability between groups themselves. Although permutation tests were conducted to address this issue, future studies with larger and more balanced samples are needed to confirm our results. Another limitation is that the majority of younger adults in our study spoke a third language. Although we focused on comparing the first two languages, the age differences in the third language profile might introduce confounding effects. Lastly, the older adults in this study had relatively lower proficiency in Mandarin and less balanced language use (i.e., lower entropy), which may not accurately represent the profiles of more balanced older multilinguals. Future studies should include age groups with more comparable language backgrounds (i.e., balanced bilinguals) to better control for the confounding effect of language proficiency on age-related differences.

In summary, the present study investigated the age-related differences in speech and brain structure from a multilingual perspective. By employing an integrated measure of speech quality, we revealed significant effects of age, language, and their interaction on speech performance. More experience of mixed use of both languages was also beneficial for speech production in both languages. Significant gray matter volume difference was found between younger and older adults. Older adults', but not younger adults' speech was associated with gray matter volume in right middle frontal and temporal gyri, and left supramarginal gyrus. The contribution of the putamen on speech in older adults was further modulated by multilingual experience. These findings underscore the complex interplay between age, multilingualism, and brain structure, providing valuable insights into the neural mechanisms underlying multilingual speech performance.

Funding

This work was supported by the Science and Technology Development Fund, Macao S.A.R (FDCT, 0153/2022/A), the National Natural Science Foundation of China (32200845), and the Start-up Research Grant (SRG2022-00003-ICI) and Multi-Year Research Grant (MYRG2022-00148-ICI) from the University of Macau to Haoyun Zhang.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Committee of the University of Macau (protocol code BSERE22-APP001-ICI, approved on 17 March 2022).

Informed Consent Statement

Informed consent was obtained from each participant prior to the beginning of the study.

CRediT authorship contribution statement

Hanxiang Yu: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Keyi Kang:** Investigation, Data curation, Conceptualization. **Christos Pliatsikas:** Writing – review & editing, Conceptualization. **Yushen Zhou:** Data curation. **Haoyun Zhang:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the staff at the Centre for Cognitive and Brain Sciences (CCBS) at the University of Macau where the study was conducted, for their support. Authors declare no competing interests.

Data availability

The data, and analysis scripts are openly available on OSF Platform athttps://osf.io/x7a4z/?view_only=2f8acbe218154faca81c1e443ab8a606. Raw imaging data are available upon request.

Reference

Abutalebi, J., Della Rosa, P.A., Green, D.W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., Costa, A., 2012. Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cerebral Cortex 22 (9), 2076–2086.
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., 2008. Control mechanisms in bilingual language production: Neural evidence from language switching studies. Language and Cognitive Processes 23 (4), 557–582.
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.
Acheson, D.J., Wells, J.B., MacDonald, M.C., 2008. New and updated tests of print exposure and reading abilities in college students. Behavior Research Methods 40 (1), 278–289.
Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.
Bilodeau-Mercure, M., Lortie, C.L., Sato, M., Guitton, M.J., Tremblay, P., 2015. The neurobiology of speech perception decline in aging. Brain Structure and Function 220, 979–997.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex 19 (12), 2767–2796.

Birdsong, D., 2006. Age and second language acquisition and processing: A selective overview. Language Learning 56 (s1), 9–49.

Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., Brennan, S.E., 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Language and Speech 44 (2), 123–147.

Brown, R., 1973. Development of the first language in the human species. American Psychologist 28 (2), 97–106. https://doi.org/10.1037/h0034209.

Burke, D.M., Shafto, M.A., 2008. Language and aging. The handbook of aging and cognition, 3rd ed. Psychology Press, pp. 373–443.

Burke, D.M., Shafto, M.A., 2011. Language and aging. The handbook of aging and cognition. Psychology Press, pp. 381–451.

Bygate, M., 1999. Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. Language teaching research 3 (3), 185–214.

Cabeza, R., 2002. Hemispheric asymmetry reduction in older adults: the HAROLD model. Psychology and Aging 17 (1), 85–100.

Chang, S.-E., Kenney, M.K., Loucks, T.M., Poletto, C.J., Ludlow, C.L., 2009. Common neural substrates support speech and non-speech vocal tract gestures. NeuroImage 47 (1), 314–325.

Cheung, H., Kemper, S., 1992. Competing complexity metrics and adults' production of complex sentences. Applied Psycholinguistics 13 (1), 53–76.

Costa, A., Calabria, M., Marne, P., Hernández, M., Juncadella, M., Gascón-Bayarri, J., Lleó, A., Ortiz-Gil, J., Ugas, L., Blesa, R., 2012. On the parallel deterioration of lexico-semantic processes in the bilinguals' two languages: Evidence from Alzheimer's disease. Neuropsychologia 50 (5), 740–753.

Dai, X., Gong, Y., Zhong, L., 1990. Factor analysis of the mainland Chinese version of the Wechsler Adult Intelligence Scale. Psychological Assessment: A Journal of Consulting and Clinical Psychology 2 (1), 31.

De Craen, A.J., Heeren, T., Gussekloo, J., 2003. Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. International Journal of Geriatric Psychiatry 18 (1), 63–66.

Diaz, M.T., Rizio, A.A., Zhuang, J., 2016. The Neural Language Systems That Support Healthy Aging: Integrating Function, Structure, and Behavior. Language and Linguistics Compass 10 (7), 314–334.

Dien, J., Franklin, M.S., Michelson, C.A., Lemen, L.C., Adams, C.L., Kiehl, K.A., 2008. fMRI characterization of the language formulation area. Brain Research 1229, 179–192. https://doi.org/10.1016/j.brainres.2008.06.107.

Duchin, S.W., Mysak, E.D., 1987. Disfluency and rate characteristics of young adult, middle-aged, and older males. Journal of Communication Disorders 20 (3), 245–257.

Ferraro, F.R., Chelminski, I., 1996. Preliminary normative data on the Geriatric Depression Scale-Short Form (GDS-SF) in a young adult sample. Journal of clinical psychology 52 (4), 443–447.

Filippetti, V.A., Allegri, R.F., 2011. Verbal fluency in Spanish-speaking children: Analysis model according to task type, clustering, and switching strategies and performance over time. The clinical neuropsychologist 25 (3), 413–436.

Fotenos, A.F., Snyder, A., Girton, L., Morris, J., Buckner, R., 2005. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64 (6), 1032–1039.

Friesen, D.C., Luo, L., Luk, G., Bialystok, E., 2015. Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. Language, cognition and neuroscience 30 (3), 238–250.

Geranmayeh, F., Brownsett, S.L., Leech, R., Beckmann, C.F., Woodhead, Z., Wise, R.J., 2012. The contribution of the inferior parietal cortex to spoken language production. Brain and Language 121 (1), 47–57.

Geranmayeh, F., Wise, R.J., Mehta, A., Leech, R., 2014. Overlapping networks engaged during spoken language production and its cognitive control. Journal of Neuroscience 34 (26), 8728–8740.

Gollan, T.H., Salmon, D.P., Montoya, R.I., da Pena, E., 2010. Accessibility of the nondominant language in picture naming: A counterintuitive effect of dementia on bilingual language production. Neuropsychologia 48 (5), 1356–1366.

Gollan, T.H., Slattery, T.J., Goldenberg, D., Van Assche, E., Duyck, W., Rayner, K., 2011. Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. Journal of Experimental Psychology: General 140 (2), 186–209.

Grundy, J.G., Anderson, J.A., Bialystok, E., 2017. Neural correlates of cognitive processing in monolinguals and bilinguals. Annals of the New York Academy of Sciences 1396 (1), 183–201.

Gullifer, J.W., Chai, X.J., Whitford, V., Pivneva, I., Baum, S., Klein, D., Titone, D., 2018. Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. Neuropsychologia 117, 123–134.

Gullifer, J.W., Kousaie, S., Gilbert, A.C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., Titone, D., 2021. Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. Applied Psycholinguistics 42 (2), 245–278. https://doi.org/10.1017/S0142716420000521.

Gullifer, J. W., & Titone, D. (2018). Compute language entropy with {languageEntropy}. Retrieved from https://github.com/jasongullifer/languageEntropy. https://doi.org/10.5281/zenodo.1403272.

Gullifer, J.W., Titone, D., 2019. Characterizing the social diversity of bilingualism using language entropy. Bilingualism: Language and cognition 23 (2), 283–294.

Gullifer, J.W., Titone, D., 2021. Bilingualism: A neurocognitive exercise in managing uncertainty. Neurobiology of language 2 (4), 464–486.

Hafkemeijer, A., Altmann-Schneider, I., de Craen, A.J., Slagboom, P.E., van der Grond, J., Rombouts, S.A., 2014. Associations between age and gray matter volume in anatomical brain networks in middle-aged to older adults. Aging cell 13 (6), 1068–1074.

Heald, S.L., Nusbaum, H.C., 2014. Speech perception as an active cognitive process. Frontiers in systems neuroscience 8, 35.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nature Reviews Neuroscience 8 (5), 393–402.

Hirshorn, E.A., Thompson-Schill, S.L., 2006. Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. Neuropsychologia 44 (12), 2547–2557.

Horton, W.S., Spieler, D.H., Shriberg, E., 2010. A corpus analysis of patterns of age-related change in conversational speech. Psychology and Aging 25 (3), 708.

Housen, A., Kuiken, F., 2009. Complexity, accuracy, and fluency in second language acquisition. Applied linguistics 30 (4), 461–473.

Indefrey, P., Levelt, W.J., 2000. The neural correlates of language production. The new cognitive neurosciences, 2nd ed. MIT press, pp. 845–865.

Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. Cognition 92 (1), 101–144.

James, L.E., Burke, D.M., Austin, A., Hulme, E., 1998. Production and perception of" verbosity" in younger and older adults. Psychology and Aging 13 (3), 355.

Kaiser, H.F., 1974. An index of factorial simplicity. Psychometrika 39 (1), 31–36.

Kang, K., Xiao, Y., Yu, H., Diaz, M.T., Zhang, H., 2023. Multilingual Language Diversity Protects Native Language Production under Different Control Demands. Brain Sciences 13 (11), 1587.

Kemper, S., Kynette, D., O'Brien, K., Rash, S., Sprott, R., 1989. Life-span changes to adults' language: Effects of memory and genre. Applied Psycholinguistics 10 (1), 49–66. https://doi.org/10.1017/S0142716400008419.

Kent, R.D., 2000. Research on speech motor control and its disorders: A review and prospective. Journal of communication disorders 33 (5), 391–428.

Kilman, L., Zekveld, A., Hällgren, M., Rönnberg, J., 2014. The influence of non-native language proficiency on speech perception performance. Frontiers in psychology 5, 651.

Korenar, M., Treffers-Daller, J., Pliatsikas, C., 2023. Brain structure adapts dynamically to highly demanding bilingual experiences: Insights from interpreters and translators. Ampersand 11, 100148. https://doi.org/10.1016/j.amper.2023.100148.

Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience 2, 249.

Lemaitre, H., Goldman, A.L., Sambataro, F., Verchinski, B.A., Meyer-Lindenberg, A., Weinberger, D.R., Mattay, V.S., 2012. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? Neurobiology of aging 33 (3), 617 e611-617. e619.

Lennon, P., 1990. Investigating Fluency in EFL: A Quantitative Approach*. Language Learning 40 (3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x.

Lerman, A., Obler, L.K., 2017. Aging in bilinguals: Normal and abnormal. Psychology of Bilingualism: The Cognitive and Emotional World of Bilinguals 189–210.

Li, P., Puls, B., Tsai, E., Zhang, F.A.N., 2014. Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. Bilingualism: Language and Cognition 17 (3), 673–680. https://doi.org/10.1017/S1366728913000606.

Macwhinney, B., 2000. The CHILDES project: tools for analyzing talk. Child Language Teaching and Therapy 8. https://doi.org/10.1177/026565909200800211.

Malcorra, B.L.C., Mota, N.B., Weissheimer, J., Schilling, L.P., Wilson, M.A., Hübner, L.C., 2021. Low speech connectedness in Alzheimer's disease is associated with poorer semantic memory performance. Journal of Alzheimer's Disease 82 (3), 905–912.

Malek, A., Hekmati, I., Amiri, S., Pirzadeh, J., Gholizadeh, H., 2013. Designing and standardization of Persian version of verbal fluency test among Iranian bilingual (Turkish-Persian) adolescents. Journal of Research in Clinical Medicine 1 (1), 32–42.

Manchon, M., Buetler, K., Colombo, F., Spierer, L., Assal, F., Annoni, J.-M., 2015. Impairment of both languages in late bilinguals with dementia of the Alzheimer type. Bilingualism: Language and cognition 18 (01), 90–100.

Marin-Marin, L., Costumero, V., Ávila, C., Pliatsikas, C., 2022. Dynamic effects of immersive bilingualism on cortical and subcortical grey matter volumes. Frontiers in Psychology 13, 886222.

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O.K., Coslett, H.B., Schwartz, M.F., 2015. Neural organization of spoken language revealed by lesion-symptom mapping. Nature communications 6.

Mortensen, L., Meyer, A.S., Humphreys, G.W., 2006. Age-related effects on speech production: A review. Language and Cognitive Processes 21 (1-3), 238–290.

Moscoso del Prado Martín, F., 2017. Vocabulary, Grammar, Sex, and Aging. Cognitive Science 41 (4), 950–975. https://doi.org/10.1111/cogs.12367.

Mota, N.B., Furtado, R., Maia, P.P., Copelli, M., Ribeiro, S., 2014. Graph analysis of dream reports is especially informative about psychosis. Scientific Reports 4 (1), 3691.

Mota, N.B., Vasconcelos, N.A., Lemos, N., Pieretti, A.C., Kinouchi, O., Cecchi, G.A., Copelli, M., Ribeiro, S., 2012. Speech graphs provide a quantitative measure of thought disorder in psychosis. PloS One 7 (4), e34928.

Mota, N.B., Weissheimer, J., Finger, I., Ribeiro, M., Malcorra, B., Hübner, L., 2023. Speech as a graph: developmental perspectives on the organization of spoken language. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.

Nanchen, G., Abutalebi, J., Assal, F., Manchon, M., Démonet, J.-F., Annoni, J.-M., 2017. Second language performances in elderly bilinguals and individuals with dementia: The role of L2 immersion. Journal of Neurolinguistics 43 (Part A), 49–58.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society 53 (4), 695–699.

Nip, I.S., Blumenfeld, H.K., 2015. Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. Journal of Speech, Language, and Hearing Research 58 (3), 653–668.

Obler, L.K., Albert, M.L., 1981. Language and aging: A neurobehavioral analysis. Aging: Communication processes and disorders 107–121.

Ortega, L., 2003. Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. Applied linguistics 24 (4), 492–518. https://doi.org/10.1093/applin/24.4.492.

Osborne, J., 2011. Fluency, complexity and informativeness in native and non-native speech. International Journal of Corpus Linguistics 16 (2), 276–298.

Papeo, L., Agostini, B., Lingnau, A., 2019. The large-scale organization of gestures and words in the middle temporal gyrus. Journal of Neuroscience 39 (30), 5966–5974.

Peramunage, D., Blumstein, S.E., Myers, E.B., Goldrick, M., Baese-Berk, M., 2011. Phonological neighborhood effects in spoken word production: An fMRI study. Journal of Cognitive Neuroscience 23 (3), 593–603.

Pfefferbaum, A., Rohlfing, T., Rosenbloom, M.J., Chu, W., Colrain, I.M., Sullivan, E.V., 2013. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85 years) measured with atlas-based parcellation of MRI. NeuroImage 65, 176–193.

Pliatsikas, C., 2020. Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. Bilingualism: Language and cognition 23 (2), 459–471.

Pobric, G., Jefferies, E., Ralph, M.A.L., 2007. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. Proceedings of the National Academy of Sciences 104 (50), 20137–20141.

Poldrack, R.A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Gabrieli, J.D., 2001. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. Journal of Cognitive Neuroscience 13 (5), 687–697.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage 10 (1), 15–35.

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.

R Core Team, R. (2013). R: A language and environment for statistical computing.

Ralph, M.A.L., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. Nature reviews neuroscience 18 (1), 42–55.

Ramanoël, S., Hoyau, E., Kauffmann, L., Renard, F., Pichat, C., Boudiaf, N., Krainik, A., Jaillard, A., Baciu, M., 2018. Gray matter volume and cognitive performance during normal aging. A voxel-based morphometry study. Frontiers in aging neuroscience 10, 235.

Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D., 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cerebral Cortex 15 (11), 1676–1689.

Resnick, S.M., Pham, D.L., Kraut, M.A., Zonderman, A.B., Davatzikos, C., 2003. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. Journal of Neuroscience 23 (8), 3295–3301.

Revelle, W., 2015. Procedures for personality and psychological research. Northwestern University, Evanston, IL.

Ridderinkhof, K.R., Van Den Wildenberg, W.P., Segalowitz, S.J., Carter, C.S., 2004. Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain and Cognition 56 (2), 129–140.

Rolls, E.T., Huang, C.-C., Lin, C.-P., Feng, J., Joliot, M., 2020. Automated anatomical labelling atlas 3. Neuroimage 206, 116189. https://doi.org/10.1016/j.neuroimage.2019.116189.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V.A., Caracciolo, V., Padilla, M., Ostrosky-Solí, F., 2000. Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. Applied Neuropsychology 7 (1), 17–24.

Segalowitz, N., Hulstijn, J., 2005. Automaticity in Bilingualism and Second Language Learning. Handbook of bilingualism: Psycholinguistic approaches. Oxford University Press, pp. 371–388.

Segalowttz, N., 2007. Access Fluidity, Attention Control, and the Acquisition of Fluency in a Second Language. TESOL Quarterly 41 (1), 181–186. http://www.jstor.org/stable/40264338.

Seghier, M.L., Price, C.J., 2009. Reading aloud boosts connectivity through the putamen. Cerebral Cortex bhp123.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell system technical journal 27 (3), 379–423.

Shriberg, E., 1999. Phonetic consequences of speech disfluency. In: Proceedings of the International Congress of Phonetic Sciences.

Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A. W., 2003. Mapping cortical change across the human life span. Nature Neuroscience 6 (3), 309–315.

Sowell, E.R., Thompson, P.M., Toga, A.W., 2004. Mapping changes in the human cortex throughout the span of life. The Neuroscientist 10 (4), 372–392.

Tachbelie, M.Y., Abate, S.T., Schultz, T., 2020. Analysis of globalphone and ethiopian languages speech corpora for multilingual asr. In: Proceedings of the Twelfth Language Resources and Evaluation Conference.

Tremblay, P., Deschamps, I., 2016. Structural brain aging and speech production: a surface-based brain morphometry study. Brain Structure and Function 221, 3275–3299.

Turker, S., Kuhnke, P., Eickhoff, S.B., Caspers, S., Hartwigsen, G., 2023. Cortical, subcortical, and cerebellar contributions to language processing: A meta-analytic review of 403 neuroimaging experiments. Psychological Bulletin.

Ullman, M.T., 2001. A neurocognitive perspective on language: The declarative/procedural model. Nature Reviews Neuroscience 2 (10), 717–726.

van den Berg, F., Brouwer, J., Tienkamp, T.B., Verhagen, J., Keijzer, M., 2022. Language entropy relates to behavioral and pupil indices of executive control in young adult bilinguals. Frontiers in Psychology 13, 1760.

Viñas-Guasch, N., Wu, Y.J., 2017. The role of the putamen in language: a meta-analytic connectivity modeling study. Brain Structure and Function 222, 3991–4004.

Visser, M., Jefferies, E., Ralph, M.L., 2010. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. Journal of Cognitive Neuroscience 22 (6), 1083–1094.

Voits, T., Pliatsikas, C., Robson, H., Rothman, J., 2020. Beyond Alzheimer's disease: Can bilingualism be a more generalized protective factor in neurodegeneration? Neuropsychologia 147, 107593. https://doi.org/10.1016/j.neuropsychologia.2020.107593.

Wechsler, D., 1997. Wechsler adult intelligence scale. Frontiers in psychology.

Wilson, S.M., Isenberg, A.L., Hickok, G., 2009. Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. Human brain mapping 30 (11), 3596–3608.

Winkler, R., Brückl, M., Sendlmeier, W., 2003. The aging voice: an acoustic, electroglottographic and perceptive analysis of male and female voices. Proc. of ICPhS 3, 2869–2872.

Wong, P.C., Ettlinger, M., Sheppard, J.P., Gunasekera, G.M., Dhar, S., 2010. Neuroanatomical characteristics and speech perception in noise in older adults. Ear and hearing 31 (4), 471–479.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology 73 (1), 3–36.

Yan, X., Kim, H.R., Kim, J.Y., 2021. Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency (CAF) features of speaking performances on the Aptis test. Language Testing 38 (4), 485–510.

Zhang, H., Wu, Y.J., Thierry, G., 2020. Bilingualism and aging: A focused neuroscientific review. Journal of Neurolinguistics 54, 100890.

Zimmerman, M.E., Brickman, A.M., Paul, R.H., Grieve, S.M., Tate, D.F., Gunstad, J., Cohen, R.A., Aloia, M.S., Williams, L.M., Clark, C.R., 2006. The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan. The American journal of geriatric psychiatry 14 (10), 823–833.