OXFORD

Neural signatures of second language proficiency in narrative processing

Ruiqing Zhang[1,2,†], Jing Wang[1,3,†], Hui Lin[2], Nicholas B. Turk-Browne[4], Qing Cai[1,3,5,*]

[1]Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), Affiliated Mental Health Center (ECNU), Institute of Brain and Education Innovation, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China,
[2]Shanghai Key Laboratory of Artificial Intelligence in Learning and Cognitive Science, LAIX Inc, Shanghai 200090, China,
[3]Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai 201210, China,
[4]Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, United States,
[5]NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai 200126, China

*Corresponding author: Qing Cai, East China Normal University, Shanghai 200062, China. Email: qcai@psy.ecnu.edu.cn
[†]R. Zhang and J. Wang contributed equally.

Making sense of speech in a second language relies on multiple abilities. Differences in brain activity related to proficiency in language tasks have often been attributed to processing demands. However, during naturalistic narrative comprehension, listeners at different proficiency levels may form different representations of the same speech. We hypothesized that the intersubject synchronization of these representations could be used to measure second-language proficiency. Using a searchlight-shared response model, we found highly proficient participants showed synchronization in regions similar to those of native speakers, including in the default mode network and the lateral prefrontal cortex. In contrast, participants with low proficiency showed more synchronization in auditory cortex and word-level semantic processing areas in the temporal lobe. Moderate proficiency showed the greatest neural diversity, suggesting lower consistency in the source of this partial proficiency. Based on these synchronization differences, we were able to classify the proficiency level or predict behavioral performance on an independent English test in held-out participants, suggesting the identified neural systems represented proficiency-sensitive information that was generalizable to other individuals. These findings suggest higher second-language proficiency leads to more native-like neural processing of naturalistic language, including in systems beyond the cognitive control network or the core language network.

Key words: narrative; shared response model (SRM); second language proficiency; speech; intersubject correlation.

Introduction

We rely on shared knowledge to understand each other, yet knowledge sharing relies on effective communication. A story "resonates with us" not only metaphorically but also physically: During perception or recall of continuous complex narratives, multiple neural systems track the transitions of events at different timescales (Baldassano et al. 2017, 2018). Neural responses to the same narrative are synchronized across multiple audience or readers (Lerner et al. 2011; Simony et al. 2016; Regev et al. 2019) and across storytellers (Silbert et al. 2014). Higher speaker-listener neural coupling predicts successful communication (Stephens et al. 2010). Shared knowledge has also been associated with synchronized event-specific response profiles across people recalling similar events (Chen et al. 2017), between encoding and recall processes (Zadbood et al. 2017), and between movie watching and speech perception (Zadbood et al. 2017; Nguyen et al. 2019). Different scales of information are represented in different neural substrates (Lerner et al. 2011) and at different orders of neural dynamics (Owen et al. 2021).

Various factors affect cross-individual synchronicity, such as attention (Regev et al. 2019) and interpretation of the materials (Nguyen et al. 2019). In the processing of speech in a second language (L2), these factors may reflect stable individual differences. When people listen to a talk in a non-native language, language proficiency affects the degree to which they can track

the speech stream, follow the topic, and grasp various details and implications. According to Common European Framework of Reference (CEFR; www.coe.int/en/web/common-european-framework-reference-languages), basic users can understand phrases related to areas of their immediate priority if the articulation is clear and slow, whereas proficient users can understand extended discourses on complex and less familiar topics with only occasional missing details.

Three lines of findings on L2 or bilingual processing are most relevant to the present study. First, consistent with the convergence hypothesis (Green 2003), L1 and L2 processing share cognitive mechanisms and neural substrates for speech perception, semantic representation, and syntactic processing (Perani et al. 1998; Illes et al. 1999; Abutalebi 2008; Kotz 2009; Sebastian et al. 2011; Costa and Sebastian-Galles 2014). Second, L2 proficiency or ability has been associated with cognitive and neural activational differences that cannot be accounted for by the age of acquisition (Chee et al. 2001; Wartenburger et al. 2003; Perani and Abutalebi 2005; Stein et al. 2009; Hamrick 2015; Hamrick et al. 2018). Third, these neural functional differences between the use of languages are often attributed to the difference in processing demand: greater efforts are required for L2 or bilingual processing, either due to the less frequent use of the language, the transfer between linguistic knowledge, or the involvement of additional cognitive control/monitoring compared with monolinguals

(Green 2003; Golestani et al. 2006; Abutalebi 2008; Luk et al. 2012; Costa and Sebastian-Galles 2014; Grant et al. 2015; Berken et al. 2016; Weber et al. 2016; Tao et al. 2021). However, it is unaddressed how the low-level understanding of language content could possibly lead to a deviation of the neural representation from those with high proficiency.

Here we further propose that during naturalistic narrative processing, variation in language proficiency leads to not only cognitive control differences, but also diverse understanding and representations of the same speech, whether or not modulated by top-down process. These diversified representations will disrupt the neural synchronizations across individuals, especially in brain areas associated with narrative processing. We hypothesize that people with a similar proficiency level will have similar neural response patterns, which differ from those at a different proficiency level. Although non-proficient learners may display local topographic discrepancies in the neural response, we hypothesize that they still rely on the same microanatomical substrates and share neural signatures due to the same levels of understanding of the second-language narratives. These shared neural signatures can be identified when the local idiosyncrasies are accounted for.

The present study uses shared response model (SRM; Chen et al. 2015; Zhang et al. 2016) to investigate the neural signatures of second language proficiency during narrative comprehension. Complex naturalistic speech materials, namely TED talks, were used to examine higher-order processes that may not be easily observed in language tasks that tap into segregated cognitive variables. Identifying cross-individual synchronization in speech processing can be challenging because fine-grained semantic encoding patterns in the brain bear considerable individual differences (e.g. Huth et al. 2016). We apply SRM to partition multi-subject fMRI data into the common neural response profiles and subject-specific local functional topographies. The approach of using SRM in a searchlight has been validated by showing optimal performance in extracting consistent neural profiles within a group of individuals (Chen et al. 2016). We expect the neural response that is specific to low-proficiency individuals to be synchronized only in earlier perceptual and word-level semantic areas, namely the auditory cortex and ventral temporal cortex, whereas high-proficiency-sensitive responses will be additionally synchronized in higher-order regions involved in narrative comprehension, such as lateral and medial prefrontal cortex, temporoparietal cortex, and posterior cingulate cortex/precuneus (Lerner et al. 2011; Honey et al. 2012; Simony et al. 2016; Chen et al. 2017; Zadbood et al. 2017; Nguyen et al. 2019). Other candidate areas include the posterior middle frontal gyrus and the precentral-postcentral areas (Honey et al. 2012). These high-level areas are expected to present proficiency-sensitive neural responses.

Materials and methods
Participants
A total of 47 healthy participants (32 females, mean age = 22.7 years, SD = 2.7) were recruited from the East China Normal University community. Participants were native Mandarin Chinese speakers with English as the second language (mean age of acquisition in listening = 8.7 years, SD = 2.7), right-handed, reported normal hearing and had no history of neurological disease or language impairment. Language history was collected using Language History Questionnaire version 2.0 (Li et al. 2014).

All participants provided informed consent that was approved by the East China Normal University Institutional Review Board.

Stimuli and experimental design
Stimuli of the main study were the audio tracks of four TED talks (www.ted.com), the topics of which were communication skills ("Talk nerdy to me" by Melissa Marshall; "Try something new for 30 days" by Matt Cutts; "The brain benefits of deep sleep – and how to get more of it" by Dan Gartenberg; and "Keep your goals to yourself" by Derek Sivers). The speeches lasted for 4.25 (averaged speed 170 words/min), 6.48 (60 words/min), 3.17 (80 words/min), and 2.98 min (82 words/min), respectively. The talks were initially selected based on the subjective criteria that the topics were familiar and relevant to college students and that the speech was clear and intelligible. Three independent raters, including one experienced English teacher, then subjectively selected the four talks out of five as being "of moderate difficulty to a Chinese college student who learns English as a second language". Post hoc examination of the scripts of the talks using Coh-Metrix (Graesser et al. 2011; https://soletlab.asu.edu/t-e-r-a/) showed the texts were of average to high narrativity and high deep cohesion (Table S1), suggesting they were easy to comprehend for the native. Participants were asked to listen to the talk with eyes open while being scanned. The four talks were presented in four separate scan runs, synchronized with MRI acquisition onset using PsychoPy (Peirce 2007). To ensure that the participants paid attention to the stimuli, participants were asked to try their best to verbally report the content of the preceding talk in their native language after each run.

Behavioral assessment of second language proficiency
Participants performed two behavioral tests before the main experiment. Phonological perception ability was measured using a phoneme discrimination task. In each trial, participants heard two word-like sounds in sequence and judged whether the two items sounded the same. The sounds were English words in the form of consonant(s)-vowel-consonant(s) pronounced by a native English speaker. Each pair was either the same or differed by one phoneme.

Due to the diversified English test history as reported by the participants, we utilized the Cambridge Assessment of English on Listening (https://www.cambridgeenglish.org/exams-and-tests/) to gage their proficiency in a common framework. The exam used is aligned with the CEFR, which grades language proficiency on a six-level scale ranging from A1 to C2. Individuals at A1 or A2 proficiency level were categorized as basic users, those at B1 or B2 as independent users, and those at C1 or C2 as proficient users (https://rm.coe.int/1680459f97). For this study, we employed examination materials corresponding to CEFR levels A2, B1, B2, and C1, whereas A1 and C2 materials were excluded, as all participants passed China's College English Test-Band 4, which is equivalent to an A2 level, and the most advanced learner in the sample earned an 8 in the IELTS test, indicative of a C1 level. We subsequently conducted a post hoc analysis of participants' exam results, which verified the adequacy of the questions in grouping them into three categories (Fig. 1). Participants listened to different paragraphs and answered questions about the materials. Performance was used to assign participants into the high-, moderate-, and low-proficiency groups, as described in the *Results* section.
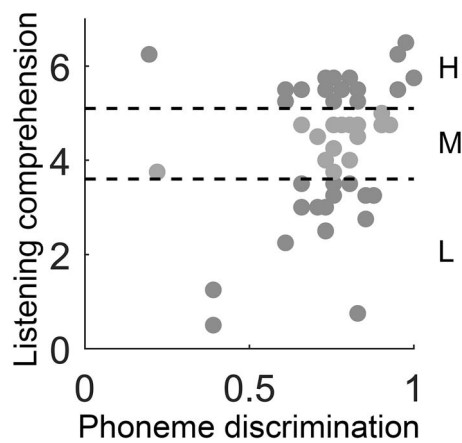
Fig. 1. Scatterplot of scores on the Cambridge listening comprehension test and accuracy of phoneme discrimination for low-, moderate-, and high-proficiency groups. Dotted lines separate proficiency groups.

MRI acquisition

Whole-brain images were collected using a 3.0-Tesla Siemens Trio MRI scanner (Siemens, Erlangen, Germany) with a 64-channel head coil. Functional images were acquired with an inter-leaved multiband echo-planar imaging sequence (TR = 1000 ms, TE = 32.6 ms, FoV = 192 mm, 2 mm isotropic voxel, 55° flip angle, 72 transversal slices, multiband factor 6). For offline correction for the susceptibility-induced distortions on EPI images, a pair of spin echo volumes was acquired in opposing phase encode directions (anterior to posterior and posterior to anterior). Pulse and respiration during task were collected using a pulse detector on a finger and a respiration detection belt wrapped around the chest. T1-weighted anatomical image was obtained using a MPRAGE sequence (TR = 2530 ms, TE = 2.34 ms, FoV = 256 mm, 7° flip angle, 192 slices). Participants wore MRI-compatible passive noise-canceling headphones.

Data analysis

Preprocessing

Images were preprocessed using FSL FEAT 5 (https://fsl.fmrib.ox. ac.uk/fsl/fslwiki/FEAT). Susceptibility-induced distortions were corrected using FSL TOPUP (Andersson et al. 2003). Off-resonance field image was derived from the two spin-echo field mapping images, with unit converted to angular frequency, and then submitted for B0 unwarping. Functional images were spatially realigned, distortion corrected, registered to the structural image and normalized to MNI152 template, temporally filtered (128 sec high-pass filter), and smoothed with a 5-mm FWHM Gaussian kernel. Respiration and pulse recordings were processed using FSL PNM (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PNM) to generate the covariates of physiological noise, which were regressed out from the preprocessed signal along with the head motion parameters.

Searchlight SRM

The subsequent analyses were perfformed using customized python scripts and the Brain Imaging Analysis Kit (Kumar et al. 2022; http://brainiak.org). In general, the SRM technique projects multi-subject data into a common, low-dimensional feature space. Given time-by-voxel matrices, SRM finds a voxel-by-feature transformation matrix for each subject and a feature-by-time matrix that is shared across all subjects. Unlike inter-subject correlation analysis that relies on the assumption of voxel-to-voxel alignment of functional topography across individuals after

structural alignment, SRM takes into account idiosyncrasies in functional topographies by estimating both the latent shared component and the subject-specific basis (Chen et al. 2015). This approach also allows for the partitioning of overall shared, group-specific, and subject-specific responses. For this study, to identify the neural signatures for different levels of proficiency, shared response modeling was applied in a leave-one-participant-out cross-validation procedure within each searchlight cube. The cube contained 27 voxels and moved by a step of one voxel. The number of shared neural features was set to 10 (Chen et al. 2015). In each cross-validation fold in each searchlight location, three SRMs, one per proficiency group, were separately trained using all but the left-out participants' data within each proficiency group. Three shared neural feature matrices (feature × time point) were learned. The feature matrices were applied separately to the left-out participant to reconstruct three sets of neural responses over time. These reconstructed responses represented three different predictions of the held-out participant's neural response based on three different hypothesized proficiency levels. Three Pearson's correlation scores were computed between the original signal and each of the mean reconstructed time series. If the reconstruction derived from the actual proficiency group of the participant presented the highest correlation, this searchlight location was considered to be proficiency-sensitive.

For each participant, the correlation scores of these proficiency-sensitive locations were transformed to Fisher' Z scores. The normalized correlation maps were averaged within each proficiency group. The maps were further thresholded against a null distribution generated by a 1000-iteration random permutation (Winkler et al. 2014). The resulting maps indicated the voxels that were reliably identified as being proficiency-sensitive across participants at the same proficiency level.

To further identify regions that were more characteristic to one proficiency group, the pairwise differences between the mean Z maps were computed between each pair of proficiency groups. These difference Z maps were thresholded against a 1000-iteration random permutation. The thresholded maps showed voxels of which the response profiles were more characteristic for one proficiency level than the other.

Classification of proficiency level

To verify the sensitivity to proficiency of the brain maps revealed by searchlight SRM, classifiers were trained on the correlation maps to identify the proficiency level of a participant. In each cross-validation fold, one participant was left out. The training exemplars were the correlation between the actual data of each participant in the training set and the time series reconstructed based on the shared response of this training participant's proficiency group. Support vector machine classifiers were trained and applied to the three correlation maps of the left-out participant to identify the group membership. A binary absolute accuracy score was determined by whether the predicted group label matched the actual, behaviorally determined label of the participant. The theoretical chance-level accuracy was 0.33. To statistically test whether the mean accuracy over participants were above chance level, the classification procedures were repeated on randomly shuffled labels of the language proficiency for 100,000 times to form a null distribution of the classification accuracy.

Prediction of second-language comprehension scores

The machine-learning procedure was the same as the classification analysis, except that a ridge regression model was trained to estimate the relation between the correlation patterns and the

English comprehension scores of the training participants. The trained model was applied to correlation pattern of the left-out participant to predict the comprehension score. Pearson's correlation between the predicted scores from all the cross-validation folds and the actual scores over participants was calculated as the metric of model performance.

Results

Behavioral results and assignment of proficiency groups

Participants were assigned to either low-proficiency (LP), moderate-proficiency (MP), or high-proficiency level (HP) group based on their Cambridge listening comprehension test scores. The mean score on Cambridge listening comprehension test over participants was 4.27 out of 6.5 (65.38%), SD = 1.38. Participants were assigned to LP if the score was 1 SD below the mean or lower and to HP if their score was 1 SD above the mean or higher, and to MP otherwise (Fig. 1). This resulted in 15 participants in LP (scores ranged from 0.5 to 3.5), 16 participants in MP (scores ranged from 3.75 to 5), and 16 participants in HP (scores ranged from 5.25 to 6.5). As a sanity check, a one-way ANOVA on scores of the three groups showed a significant group effect ($F = 84.32$, $P < 0.0001$) and post hoc Wilcoxon tests between pairs of groups indicated significant differences between HP and MP ($P < 10^{-6}$) and between MP and LP ($P < 10^{-5}$).

The mean phoneme discrimination accuracy of HP, MP, and LP were 0.82 (SD = 0.12), 0.75 (SD = 0.16), and 0.71 (SD = 0.15). ANOVA revealed no significant effect of groups ($F = 2.132$, $P = 0.131$; Fig. 1), suggesting similar lower-level abilities across groups. For HP, MP, and LP groups, respectively, the self-reported average years of experience with English were 16, 13, and 13 years. The average hours of English use were 6, 3, and 2 h per day. The self-rated ability of English listening (1 = very unproficient, 7 = very proficient) was 5, 4, and 3.

Linear regressions using phoneme discrimination to predict comprehension scores showed that no significant relationship existed in HP or MP ($ps > 0.05$). In the LP group, the comprehension scores were found positively predicted by phoneme discrimination ability ($R^2 = 0.29$, $P = 0.022$), suggesting that overall narrative comprehension ability at the low end may depend on lower-level phonological processes.

Neural signatures of second language proficiency

By using the SRM in a searchlight procedure, we discovered neural synchronization across individuals at the same level of English proficiency during speech comprehension throughout the brain. For a given searchlight, if the response profile of one participant can be most accurately reconstructed based on the shared response of other participants at the same proficiency group, this region was considered characteristic to a proficiency level. Consistent responses for the high proficiency (HP) group were found throughout the bilateral medial and lateral frontal and occipital cortices, postcentral and precentral gyri, anterior cingulate cortices (ACC), posterior cingulate cortex/precuneus (PCC), insula, parahippocampal gyri (PHG), left intraparietal sulcus (IPS), and right supramarginal gyrus (SMG; Fig. 2A). Compared with the moderate proficiency (MP) group, the HP group showed stronger synchronization in several clusters in the left anterior to middle temporal lobe (ATL), bilateral superior and middle frontal gyri, postcentral gyri, middle occipital gyri (MOG), right PHG, and right insula. Compared with the low proficiency (LP) group, the HP group showed stronger synchronization in bilateral lingual gyri,

MOG, and right inferior frontal gyrus (IFG). In short, the HP group showed widespread synchronization in default mode network areas and in proposition representation areas.

For the MP group, consistent responses were found in the bilateral ATL, auditory cortices, posterior superior temporal sulci (pSTS), IPS, postcentral gyri, orbitofrontal cortex, ACC, PCC, PHG, left SMG, MOG, right IFG, and lingual gyrus (Fig. 2B). No regions showed greater synchronization for the MP group compared against either the HP or LP groups.

For the LP group, consistent responses were found in the bilateral auditory cortices, middle to posterior middle temporal gyri, IFG, precentral and postcentral gyri, ACC, PCC, superior parietal lobule (SPL), right insula, and lingual gyrus (Fig. 2C). No regions showed greater synchronization compared with the HP group. The LP group showed greater synchronization in the left insula, Rolandic operculum, STS, bilateral precentral gyri, mPFC, PCC, right PHG, and SPL.

Predicting the proficiency level of individual participants

The mean accuracy of three-way classification that identified the proficiency group membership of individual participants was 0.49 (mean chance level at 0.33; $P < 0.05$; Fig. 3A. The predicted English comprehension scores were also significantly correlated with the true comprehension scores over participants, $r = 0.47$, $P = 0.008$ (Fig. 3B).

Discussion

The present study found that inter-subject neural synchronization during second language narrative comprehension was broadly tuned by language proficiency. By using naturalistic continuous speech as the probe and searchlight SRM to align multi-subject responses with tolerance on local spatial idiosyncrasies, we identified proficiency-specific neural profiles that were more widely distributed than was previously thought. The broadly distributed informative voxels suggest that much of the neural system was tuned to the naturalistic stimuli. Such synchronization could be identified using SRM once the individual differences in functional topographies are accounted for.

The high proficiency group showed synchronization in regions that were similar to native speakers. Shared response profiles or activations across individuals during narrative processing in one's native language have been repeatedly observed in the bilateral lateral frontal areas, mPFC, ACC, PCC, SMG, IPS, and ventral temporal areas (Mazoyer et al. 1993; Lerner et al. 2011; Honey et al. 2012; Simony et al. 2016; Chen et al. 2017; Yeshurun et al. 2017; Zadbood et al. 2017; Nguyen et al. 2019), many of which compose the default mode network (Raichle et al. 2001). The default mode network has been implicated in tracking information over longer timescales (Lerner et al. 2011; Honey et al. 2012) and representing high-level semantics (Chen et al. 2017; Yeshurun et al. 2017; Nguyen et al. 2019). A closer investigation has suggested that the PCC/precuneus is also engaged in higher-order information structures regardless of the semantic congruency, as long as the structure is consistent over time (Aly et al. 2018). Whereas synchronization in the PCC/precuneus was present in all groups, the proficient participants showed a particularly widespread synchronization in the medial and lateral frontal areas. The dorsomedial prefrontal cortex is one of the most robustly identified areas in various semantic processing tasks (Binder et al. 2009). Strong engagement of dorsomedial prefrontal cortex has been found in explicit phrasal composition (Graessner et al. 2021).
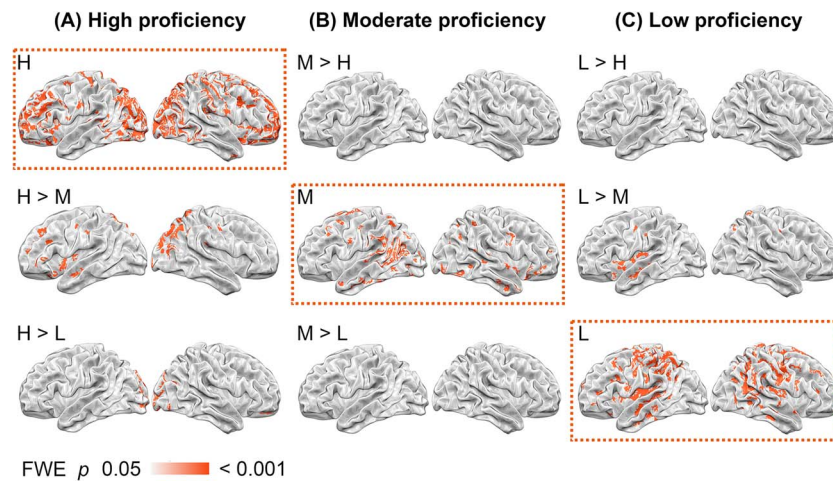
Fig. 2. Proficiency-specific shared responses. Searchlight maps of neural synchronization across individuals for high- (A), moderate- (B), and low-(C) proficiency groups. The three diagonal boxes show the regions with proficiency-specific responses for each group, namely where a participant's neural responses were most accurately reconstructed from the shared responses of other participants in the same group. The off-diagonal maps show where this same-group synchronization differed from the same-group synchronization of other groups. The brain regions where the reconstructed signals were significantly correlated with the actual signatures, whether or not being proficiency-sensitive, were widely distributed in the frontal, temporal, parietal, and occipital cortices (Fig. S1).
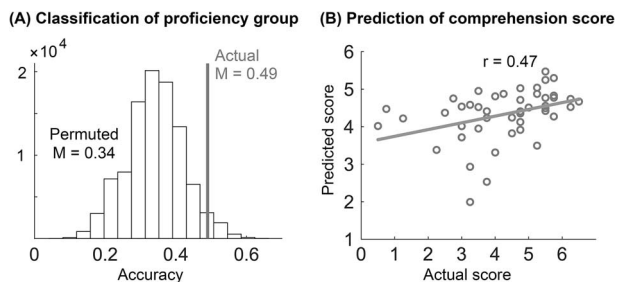

Fig. 3. Performance of proficiency level prediction. (A) Classification accuracy for predicting proficiency group from the synchronization maps of individual participants. The histogram shows the distribution of classification accuracies under null hypothesis of scrambling group membership. The vertical line shows the actual mean classification accuracy over participants. (B) Scatter plot of the actual and predicted scores in the Cambridge listening comprehension test.

The ventromedial prefrontal cortex is critical in integrating new information with prior schema (van Kesteren et al. 2010) and has been proposed to contribute to the very late stage of semantic composition that connects the core semantic systems with social cognition and episodic memory (Pylkkänen 2019; Jackson et al. 2020). Thus, these results suggest that proficient L2 users rely on native-like neural systems during L2 speech processing, which incorporate regions beyond cognitive control areas or the core language network. Future study is required to compare the neural response profiles of native speaker and second-language learners to the same stimuli to characterize the neural signatures of the ultimate attainment of a language.

Previous studies have associated the improved L2 proficiency with reduced lateral prefrontal activations (Chee et al. 2001; Stein et al. 2009) and increased ACC-lateral frontal connectivity (Grant et al. 2015) during lexical-semantic processing. The present results showed that the activation profiles to continuous speech input in the dorsolateral prefrontal cortex, a typically observed area in selective attention and working memory tasks (Curtis and D'Esposito 2003), were more alike among proficient individuals than the less proficient. These findings jointly suggest

the speculation that for proficient L2 users, enhanced cognitive control from medial frontal cortex may increase local efficiency in lateral areas (decreased activation magnitude) and entrainment of lateral responses with the external narrative, increasing cross-participant synchronization.

The low proficiency group also showed considerable within-group consistency, mostly in the temporal and temporoparietal areas. The synchronized areas across these participants overlapped with phonological processing and the ventral lexical-semantic pathway proposed in the dual-stream model of speech perception (Hickok and Poeppel 2007). Note that lower proficiency individuals showed statistically comparable ability in English phoneme discrimination as other groups, but only this group exhibited an association between behavioral performance on phonological and comprehension tasks. This suggests that participants with lower proficiency managed to perceptually track the continuous input and word-level semantics, but the narrative-level comprehension was disrupted or at least insufficient to reach agreement. Although we expected less proficient language learners to display synchronization in areas associated with phonological processing or word semantics, we did not anticipate that this synchronization would be exclusive to the low proficiency level. We found that the response of a low-proficiency individual could be better reconstructed based on other low-proficiency individuals than on proficient individuals. One possible explanation is based on the findings that attention modulates the response of early sensory cortices and enhances inter-subject synchronization in both early and higher-order areas (Regev et al. 2019). Perhaps low-proficiency participants are only able to attend to information represented at small time scales, whereas more proficient participants are less focused on phonological or word-level information, but instead pay more attention to information at the phrase, sentence, and paragraph levels.

The moderate proficiency group showed the least internal consistency among the three groups, which coincides with findings that moderately proficient L2 can be associated with more deviations from the typical or L1-like activations (Dehaene et al. 1997; Mouthon et al. 2020). Interestingly, the low proficiency group

showed greater synchronization than the moderate proficiency group in the temporal cortices as well as in the ACC and PCC. Considering that the participants were assigned to discrete groups based on their comprehension scores, originally a continuous measure, we speculate that the reduced consistency of moderate participants reflects greater heterogeneity in this group. If there is a state of ultimate attainment for L2 acquisition, a moderately proficient learner can be someone undergoing rapid development of skills, someone encountering a temporary bottleneck, or someone getting fossilized. Various types of intermediate states add to the diversity of the group compared with the experts or beginners.

A seeming surprising finding is the systematic response of the occipital lobe to speech processing. In fact, synchronized responses in lingual gyrus were seen in all proficiency groups. The involvement of lingual gyrus in sentence-level speech processing has been reported in various studies (Rodd et al. 2005; Zekveld et al. 2006; Brennan et al. 2012; Hasson et al. 2018). Indeed, there are several findings consistent with a role for occipital cortex in language proficiency. Congenitally blind individuals recruit occipital cortex in auditory sentence comprehension task (Bedny et al. 2011). Attention to speech input allows the content-specific responses to spread to the visual cortices (Regev et al. 2019). Higher-order correlations within visual areas are able to characterize content-specific inter-subject consistency during story listening (Owen et al. 2021). Cortical thickness of the lingual gyrus covaries with temporal cortex, inferior parietal cortex and inferior frontal gyrus (Chen et al. 2008). These findings are all consistent with the idea that early sensory cortices are modulated by or coactivate with higher-order areas, and may suggest that complex narrative processing is associated with information propagated across the brain. With that said, future studies with more experimentally controlled stimuli and encoding models are required to specify the underlying mechanisms that result in the proficiency-sensitive neural signatures during naturalistic processing.

Regarding the group differences in shared responses, the present results showed a paradoxical picture. Comparison of the high-proficiency-sensitive areas and the low-proficiency-sensitive areas (the "H" and "L" maps in the diagonal of Fig. 2) suggested multiple areas in the frontal and parietal cortices had different levels of within-group synchronicity, which aligned with the expectation. However, a direct comparison of the correlation between high and low proficiency groups did not reveal that many areas (the "H > L" map in Fig. 2). In this sense, the results contradicted our expectation. The reason for this seemingly inconsistent result was the different criteria for identifying proficiency-sensitive voxels. For instance, some voxel in the left MFG was shown in the H map but not the L map (both maps were on the diagonal of Fig. 2). For a low-proficiency individual, the reconstructed signals based on the HP model and the LP model might be equally similar to the actual "LP signal" (actual-to-reconstructed similarity, $r_{LH} = r_{LL}$). The voxel was thus not LP-sensitive. For a high-proficiency individual, the HP-model-based reconstruction might be more similar to the actual HP signal than the LP-model-based reconstruction ($r_{HH} > r_{HL}$). This voxel was thus HP-sensitive. In the meantime, the r_{LL} might be high enough to be indistinguishable from r_{HH}, and therefore this voxel was not shown in the H > P contrast. The missing of the frontal and parietal cortices in the H > P contrast indicated that the neural signature shared by the low-proficiency group was unable to recover the neural responses of the high-proficiency individuals.

Beyond descriptive findings, reliable accuracies in predicting the proficiency of individual participants confirmed that the neural signatures did reflect one's L2 proficiency. Instead of using distilled and sometimes unnatural experimental comparisons based on postulated cognitive constructs, we examined neural response profiles in a naturalistic language comprehension task. This enabled a more detailed neural characterization at a finer-grained proficiency level. Moreover, using searchlight SRM allowed us to identify where in the brain there were consistent responses across participants without being restricted to a rigid voxel-to-voxel correspondence. This study has illustrated a promising application of the SRM technique, namely to identify individual differences via within-group commonalities and between-group differences in neural response profile.

Acknowledgments

Author contributions

Ruiqing Zhang (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review and editing), Jing Wang (Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Software, Validation, Visualization, Writing—original draft, Writing—review and editing), Hui Lin (Conceptualization, Resources, Supervision), Nicholas Turk-Browne (Conceptualization, Formal analysis, Methodology, Software, Supervision, Validation, Writing—original draft, Writing—review and editing), Qing Cai (Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing—original draft, Writing—review and editing).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

Data availability

Data are available from the corresponding author Qing Cai (qcai@psy.ecnu.edu.cn) upon request.

References

Abutalebi J. Neural aspects of second language representation and language control. *Acta Psychol.* 2008:128(3):466–478.

Andersson JL, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage.* 2003:20(2):870–888.

Aly M, Chen J, Turk-Browne NB, Hasson U. Learning naturalistic temporal structure in the posterior medial network. *J Cogn Neurosci*. 2018:30(9):1345–1365.

Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA. Discovering event structure in continuous narrative perception and memory. *Neuron*. 2017:95(3):709–721.e5.

Baldassano C, Hasson U, Norman KA. Representation of real-world event schemas during narrative perception. *J Neurosci*. 2018:38(45):9689–9699.

Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R. Language processing in the occipital cortex of congenitally blind adults. *Proc Natl Acad Sci USA*. 2011:108(11):4429–4434.

Berken JA, Chai X, Chen JK, Gracco VL, Klein D. Effects of early and late bilingualism on resting-state functional connectivity. *J Neurosci*. 2016:36(4):1165–1172.

Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex*. 2009:19(12):2767–2796.

Brennan J, Nir Y, Hasson U, Malach R, Heeger DJ, Pylkkanen L. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang*. 2012:120(2):163–173.

Chee MWL, Hon N, Lee HL, Soon CS. Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *NeuroImage*. 2001:13(6):1155–1163.

Chen ZJ, He Y, Rosa P, Germann J, Evans AC. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb Cortex*. 2008:18(10):2374–2381.

Chen PH, Chen J, Yeshurun Y, Hasson U, Haxby JV, Ramadge PJ. A reduced-dimension fMRI shared response model. *Adv Neural Inf Process Syst*. 2015, Nips 2015:28:28.

Chen J, Honey CJ, Simony E, Arcaro MJ, Norman KA, Hasson U. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cereb Cortex*. 2016:26(8):3428–3441.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci*. 2017:20(1):115–125.

Costa A, Sebastian-Galles N. How does the bilingual experience sculpt the brain? *Nat Rev Neurosci*. 2014:15(5):336–345.

Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci*. 2003:7(9):415–423.

Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, vande Moortele PF, Lehericy S, LeBihan D. Anatomical variability in the cortical representation of first and second language. *Neuroreport*. 1997:8(17):3809–3815.

Golestani N, Alario FX, Meriaux S, Le Bihan D, Dehaene S, Pallier C. Syntax production in bilinguals. *Neuropsychologia*. 2006:44(7): 1029–1040.

Graesser AC, McNamara DS, Kulikowich J. Coh-Metrix: providing multilevel analyses of text characteristics. *Educ Res*. 2011:40(5): 223–234.

Graessner A, Zaccarella E, Hartwigsen G. Differential contributions of left-hemispheric language regions to basic semantic composition. *Brain Struct Funct*. 2021:226(2):501–518.

Grant AM, Fang SY, Li P. Second language lexical development and cognitive control: a longitudinal fMRI study. *Brain Lang*. 2015:144: 35–47.

Green D. The neural basis of the lexicon and the grammar in L2 acquisition: the convergence hypothesis. In: van, Hout R, Hulk A, Kuiken F, Towell R, editors. *The lexicon-syntax Interface in Secound language acquisition*. Amsterdam: John Benjamins; 2003.

Hamrick P. Declarative and procedural memory abilities as individual differences in incidental language learning. *Learn Individ Differ*. 2015:44:9–15.

Hamrick P, Lum JAG, Ullman MT. Child first language and adult second language are both tied to general-purpose learning systems. *Proc Natl Acad Sci USA*. 2018:115(7):1487–1492.

Hasson U, Egidi G, Marelli M, Willems RM. Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. *Cognition*. 2018:180:135–157.

Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007:8(5):393–402.

Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*. 2012:76(3):668.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*. 2016:532(7600):453–458.

Illes J, Francis WS, Desmond JE, Gabrieli JDE, Glover GH, Poldrack R, Lee CJ, Wagner AD. Convergent cortical representation of semantic processing in bilinguals. *Brain Lang*. 1999:70(3): 347–363.

Jackson RL, Bajada CJ, Lambon Ralph MA, Cloutman LL. The graded change in connectivity across the ventromedial prefrontal cortex reveals distinct subregions. *Cereb Cortex*. 2020:30(1): 165–180.

Kotz S. A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang*. 2009:109(2-3):68–74.

Kumar M, Anderson MJ, Antony JW, Baldassano C, Brooks PP, Cai MB, Chen P-HC, Ellis CT, Henselman-Petrusek G, Huberdeau D, et al. BrainIAK: the brain imaging analysis kit. *Aperture Neuro*. 2022:1(4):19.

Lerner Y, Honey CJ, Silbert LJ, Hasson U. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci*. 2011:31(8):2906–2915.

Li P, Zhang F, Tsai E, Puls B. Language History Questionnaire (LHQ 2.0): a new dynamic web-based research tool. *Biling Lang Cogn*. 2014:17(3):673–680.

Luk G, Green DW, Abutalebi J, Grady C. Cognitive control for language switching in bilinguals: a quantitative meta-analysis of functional neuroimaging studies. *Lang Cogn Process*. 2012:27(10): 1479–1488.

Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J. The cortical representation of speech. *J Cogn Neurosci*. 1993:5(4):467–479.

Mouthon M, Khateb A, Lazeyras F, Pegna AJ, Lee-Jahnke H, Lehr C, Annoni JM. Second-language proficiency modulates the brain language control network in bilingual translators: an event-related fMRI study. *Biling Cogn*. 2020:23(2):251–264.

Nguyen M, Vanderwal T, Hasson U. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*. 2019:184:161–170.

Owen LLW, Chang TH, Manning JR. High-level cognition during story listening is reflected in high-order dynamic correlations in neural activity patterns. *Nat Commun*. 2021:12(1):5728.

Peirce JW. PsychoPy—psychophysics software in python. *J Neurosci Methods*. 2007:162(1-2):8–13.

Perani D, Abutalebi J. The neural basis of first and second language processing. *Curr Opin Neurobiol*. 2005:15(2):202–206.

Perani D, Paulesu E, Galles NS, Dupoux E, Dehaene S, Bettinardi V, Cappa SF, Fazio F, Mehler J. The bilingual brain. Proficiency and

age of acquisition of the second language. *Brain*. 1998:121(10): 1841–1852.

Pylkkänen L. The neural basis of combinatory syntax and semantics. *Science*. 2019:366(6461):62–66.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci USA*. 2001:98(2):676–682.

Regev M, Simony E, Lee K, Tan KM, Chen J, Hasson U. Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cereb Cortex*. 2019:29(10):4017–4034.

Rodd JM, Davis MH, Johnsrude IS. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex*. 2005:15(8):1261–1269.

Sebastian R, Laird AR, Kiran S. Meta-analysis of the neural representation of first language and second language. *Appl Psycholinguist*. 2011:32(4):799–819.

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci USA*. 2014:111(43): E4687–E4696.

Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun*. 2016:7(1):12141.

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

Stephens GJ, Silbert LJ, Hasson U. Speaker–listener neural coupling underlies successful communication. *Proc Natl Acad Sci*. 2010:107(32):14425–14430.

Tao L, Wang G, Zhu M, Cai Q. Bilingualism and domain-general cognitive functions from a neural perspective: a systematic review. *Neurosci Biobehav Rev*. 2021:125:264–295.

van Kesteren MTR, Fernandez G, Norris DG, Hermans EJ. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci USA*. 2010:107(16):7550–7555.

Wartenburger I, Heekeren HR, Abutalebi J, Cappa SF, Villringer A, Perani D. Early setting of grammatical processing in the bilingual brain. *Neuron*. 2003:37(1):159–170.

Weber K, Luther L, Indefrey P, Hagoort P. Overlap and differences in brain networks underlying the processing of complex sentence structures in second language users compared with native speakers. *Brain Connect*. 2016:6(4):345–355.

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. *NeuroImage*, 2014:92:381–397.

Yeshurun Y, Nguyen M, Hasson U. Amplification of local changes along the timescale processing hierarchy. *Proc Natl Acad Sci USA*. 2017:114(35):9475–9480.

Zadbood A, Chen J, Leong YC, Norman KA, Hasson U. How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb Cortex*. 2017:27(10): 4988–5000.

Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R. Top-down and bottom-up processes in speech comprehension. *NeuroImage*. 2006:32(4):1826–1836.

Zhang H, Chen P-H, Chen J, Zhu X, Turek JS, Willke TL, Hasson U, Ramadge PJ. A searchlight factor model approach for locating shared information in multi-subject fMRI analysis. 2016: arXiv e-prints. arXiv:1609.09432.