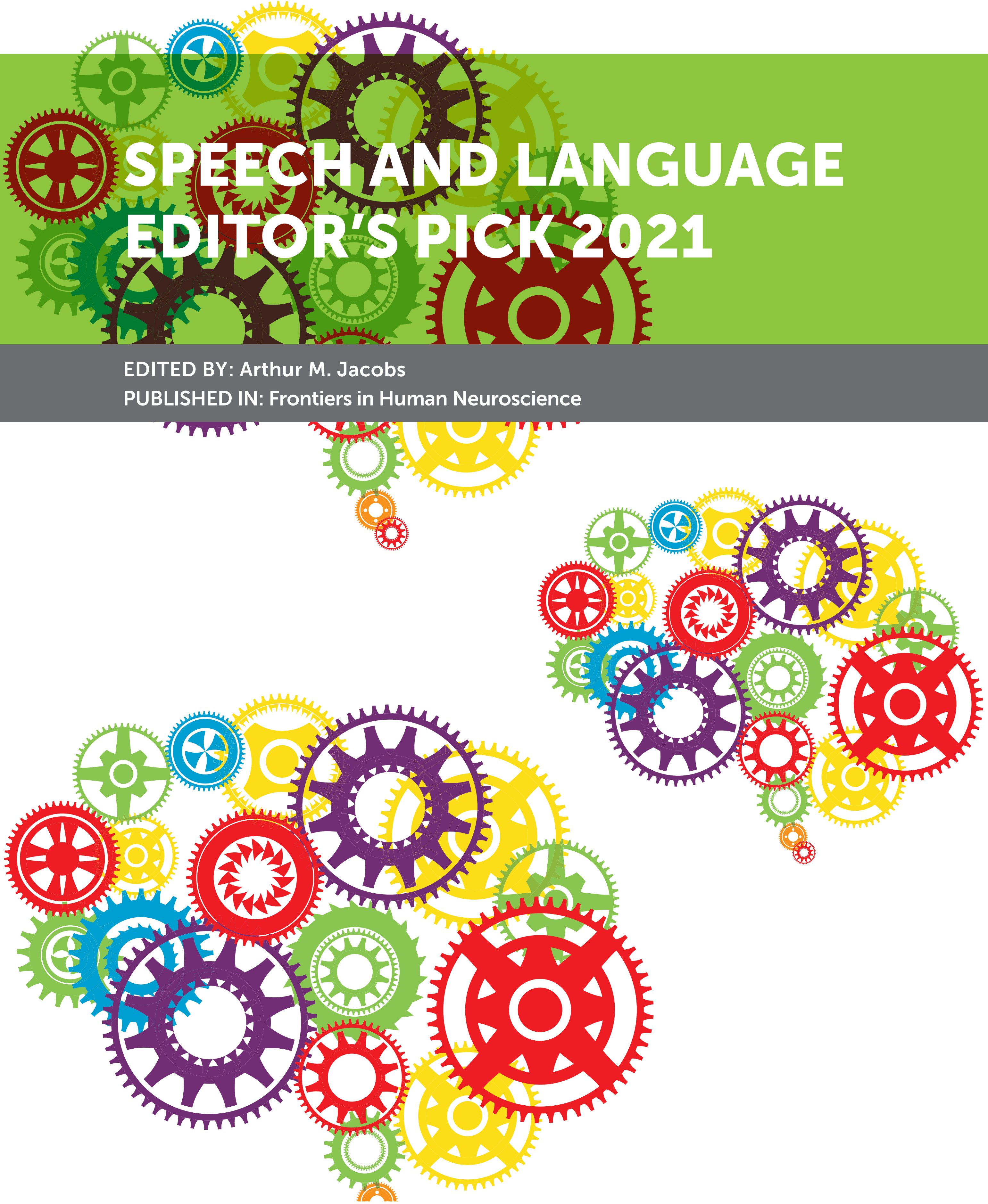
SPEECH AND LANGUAGE EDITOR'S PICK 2021



EDITED BY: Arthur M. Jacobs
PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-148-2
DOI 10.3389/978-2-88971-148-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SPEECH AND LANGUAGE EDITOR’S PICK 2021

Topic Editor:
Arthur M. Jacobs, Freie Universität Berlin, Germany

Citation: Jacobs, A. M., ed. (2021). Speech and Language Editor’s Pick 2021. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-148-2

Table of Contents

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 13 February 2019
doi: 10.3389/fnhum.2019.00027

Neural Connectivity in Syntactic Movement Processing

Eduardo Europa[1]*, *Darren R. Gitelman*[2,3,4], *Swathi Kiran*[5] *and Cynthia K. Thompson*[1,4,6]

[1] Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, [2] Advocate Lutheran General Hospital, Park Ridge, IL, United States, [3] Department of Medicine, Rosalind Franklin University of Medicine and Science, North Chicago, IL, United States, [4] The Ken and Ruth Davee Department of Neurology Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, [5] College of Health & Rehabilitation Sciences, Boston University, Boston, MA, United States, [6] Mesulam Cognitive Neurology and Alzheimer's Disease Center, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Linguistic theory suggests non-canonical sentences subvert the dominant *agent-verb-theme* order in English via displacement of sentence constituents to argument (NP-movement) or non-argument positions (wh-movement). Both processes have been associated with the left inferior frontal gyrus and posterior superior temporal gyrus, but differences in neural activity and connectivity between movement types have not been investigated. In the current study, functional magnetic resonance imaging data were acquired from 21 adult participants during an auditory sentence-picture verification task using passive and active sentences contrasted to isolate NP-movement, and object- and subject-cleft sentences contrasted to isolate wh-movement. Then, functional magnetic resonance imaging data from regions common to both movement types were entered into a dynamic causal modeling analysis to examine effective connectivity for wh-movement and NP-movement. Results showed greater left inferior frontal gyrus activation for *Wh > NP-movement*, but no activation for *NP > Wh-movement*. Both types of movement elicited activity in the opercular part of the left inferior frontal gyrus, left posterior superior temporal gyrus, and left medial superior frontal gyrus. The dynamic causal modeling analyses indicated that neither movement type significantly modulated the connection from the left inferior frontal gyrus to the left posterior superior temporal gyrus, nor vice-versa, suggesting no connectivity differences between wh- and NP-movement. These findings support the idea that increased complexity of wh-structures, compared to sentences with NP-movement, requires greater engagement of cognitive resources via increased neural activity in the left inferior frontal gyrus, but both movement types engage similar neural networks.

Keywords: syntactic movement, non-canonical sentences, sentence comprehension, functional magnetic resonance imaging, dynamic causal modeling

Edited by:
Rik Vandenberghe,
KU Leuven, Belgium

Reviewed by:
Roel M. Willems,
Radboud University Nijmegen, Netherlands
Gesa Hartwigsen,
Max Planck Institute for Human Cognitive and Brain Sciences, Germany

****Correspondence:***
Eduardo Europa
europa@u.northwestern.edu

Received: *08 October 2018*
Accepted: *21 January 2019*
Published: *13 February 2019*

Citation:
Europa E, Gitelman DR, Kiran S and Thompson CK (2019) Neural Connectivity in Syntactic Movement Processing. Front. Hum. Neurosci. 13:27. doi: 10.3389/fnhum.2019.00027

INTRODUCTION

Auditory sentence comprehension requires the rapid integration of phonological, semantic, and syntactic information and is primarily supported by a network of regions in the left perisylvian cortex. Neurocognitive models suggest that the left inferior frontal and posterior temporal areas are integral for sentence processing, but their functions and neural dynamics are not clearly understood. Sentence comprehension is affected by syntactic structure, in that canonical forms that follow the basic word order of a particular language (e.g., subject-verb-object (SVO) in English), as in (1) below, are easier to understand than non-canonical forms, as in (2) and (3), that deviate from canonical order. Further, there are several types of non-canonical sentences that engage unique linguistic processes that may engage differential neural networks.

(1) The woman weighed the boy (Active sentence; canonical).
(2) *The boy*$_i$ was weighed *(t_i)* by the woman (Passive sentence; non-canonical; NP-movement).
(3) It was *the boy*$_j$ *who*$_{i,j}$ the woman weighed *(t_i)*. (Object-cleft sentence; non-canonical; Wh-movement).

Based on linguistic theory (i.e., Government and Binding Theory Chomsky, 1986, 1995) (2) passive, and (3) object-cleft structures involve differing movement operations: NP- and wh-movement, respectively. NP-movement refers to noun phrase movement, whereas, wh-movement refers to movement of a wh-operator (e.g., *who*). In both structures, the moved constituent originates in the object position, assigned a theme by the verb, and once moved, a trace (*t*) is left behind marking its original position. In (2) the displaced theme occupies an argument (i.e., the subject) position in the sentence. However, in (3) the theme moves to a non-argument position. In both movement types, the displaced element has a dependency relationship with the trace (as noted by the subscript *i*). In addition, because object-clefts involve an embedded clause, a co-referential relation between the moved element and the head noun of the relative clause is required (denoted by the subscript *j*). This additional dependency renders the wh-movement structure in (3) more complex than the NP-movement structure in (2).

Psycholinguistic studies have examined whether these representational descriptions are associated with measurable cognitive processing costs. Cross-modal priming tasks and visual world eyetracking studies have shown increased processing time at the trace site while listening to NP- and wh-movement structures (Nagel et al., 1994; Lee, 2004; Dickey et al., 2007; Dickey and Thompson, 2009). Findings from individuals with agrammatic aphasia suggest that the double dependency in wh-movement engenders greater processing resources (Mauner et al., 1993; Dickey and Thompson, 2004; Salis and Edwards, 2005).

Functional imaging studies also have investigated the neural mechanisms of wh- and NP-movement sentences, though none have made any direct comparisons. Studies of wh-movement commonly reported activation in and around the left inferior frontal gyrus (IFG) and temporoparietal junction (TPJ) (e.g., Caplan et al., 1999, 2008; Ben-Shachar et al., 2003, 2004; Thompson et al., 2010b; Bornkessel-Schlesewsky and Schlesewsky, 2013). Not only are these regions involved with processing complex verb argument structure (Ben-Shachar et al., 2003; Thompson et al., 2010a), but they also are engaged for integrating semantic and syntactic information. Cross-linguistic studies have reported similar regions. A Hebrew study by Shetreet and Friedmann (2014) found that, when directly compared to verb movement, wh-movement elicited activity in the left IFG (BA 44/45), left posterior temporal cortex (BA 22), and medial superior frontal gyrus. Similarly, a study by Makuuchi et al. (2012) found that activity in the left pars opercularis of the IFG (BA 44) was positively correlated with distance in German sentences with scrambling, but not in those with wh-movement. Taken together, these results are in line with theories that the left posterior IFG may be involved with processes that occur after initial phrase structure building and semantic interpretation via syntactic working memory processes (Caplan et al., 1999, 2008) which precede thematic role re-analysis in the left TPJ.

Fewer studies have examined the neural correlates of NP-movement. A study comparing passive sentences to active sentences linked neural activity in the left pars opercularis and triangularis of the IFG to NP-movement and/or non-canonical verb-argument structure mapping, and activity in the left posterior middle temporal gyrus and superior parietal lobule to thematic mapping and re-analysis processes (Mack et al., 2013). These results are in line with several Japanese and Chinese NP-movement studies that also compared passive and active structures. However, it should be noted that unlike Japanese and English passives which are marked by an additional morpheme in the verb, the Chinese language has no morphological inflections (Yokoyama et al., 2007; Feng et al., 2015). Temporoparietal activity was reported in the left posterior superior temporal gyrus (STGp) (Kinno et al., 2008; Hirotani et al., 2011) for Japanese and Chinese passives, and the left superior parietal lobule (Yokoyama et al., 2006) and left inferior parietal lobule (IPL) (Yokoyama et al., 2007) for Japanese passives only–areas which have all been previously implicated in thematic re-analysis and verb-argument integration (Thompson and Meltzer-Asscher, 2014). This interpretation is further supported by an Italian study that demonstrated improved accuracy on the comprehension of passive sentences after transcranial magnetic stimulation in the left posterior parietal cortex (Finocchiaro et al., 2015). NP-movement was also associated with the left pars triangularis of the IFG for both Japanese (Yokoyama et al., 2006; Hirotani et al., 2011) and Chinese passives (Ye and Zhou, 2009), and the left pars orbitalis of the IFG in Chinese passives (Feng et al., 2015). In contrast, a few studies comparing passives to actives reported activation only in non-traditional language areas such as the left frontal operculum, caudal to the IFG (Yokoyama et al., 2007), and the postcentral gyrus (Matchin and Hickok, 2016). In summary, converging evidence across methods and languages provide support for the neural instantiation of syntactic movement which may be supported by a left hemisphere network including the TPJ and the IFG.

Two neurocognitive models of auditory sentence comprehension offer different predictions for how syntactic movement might be processed. The model by Friederici (2012) proposes that initial and higher-order syntactic processes elicit neural activity along temporo-frontal ventral and fronto-temporal dorsal pathways, respectively. The ventral tract consists of the extreme capsule fiber system and the uncinate fasciculus, and is associated with retrieval of lexical-semantic information in the middle temporal gyrus, followed by first-pass syntactic and semantic parsing in the anterior temporal lobe and anterior IFG. The dorsal tract includes the superior longitudinal/arcuate fasciculus which is involved in processing syntactic complexity. In this model, non-canonical sentences with syntactic movement first undergo phrase structure building in the left IFG, then thematic role re-analysis in the left TPJ (also see Thompson and Meltzer-Asscher (2014) for a similar model for processing verb argument structure).

On the contrary, the model proposed by Bornkessel-Schlesewsky and Schlesewsky (2013) suggests that all sentences, regardless of complexity, begin with lexical processing in the left posterior superior temporal cortex followed by temporo-frontal ventral and dorsal projections to the left frontal cortex for integration of linguistic information. In their model, the ventral tract is engaged by combinatorial semantic processes, while the dorsal tract subserves the identification and parsing of syntactic relations. These two tracts converge on the left IFG where semantics and syntax are integrated. However, this model makes two controversial claims: First, the authors argue that there are no specialized mechanisms for syntactic complexity, as they claim there is no cross-linguistic operational definition that differentiates simple from complex syntax across studies (Schlesewsky and Bornkessel-Schlesewsky, 2013). Second, they argue that the IFG is not directly involved in linguistic processing, adding to the extensive debate over the role of the left IFG in language-specific vs. domain-general functions (Hagoort, 2005; Costafreda et al., 2006; Grodzinsky and Santi, 2008; Rogalsky and Hickok, 2011). These authors associate the IFG with cognitive control and/or conflict resolution. According to their model, non-canonical sentences with syntactic movement are processed in a similar manner to canonical sentences: both engage left dorsal and ventral temporo-frontal perisylvian pathways from the posterior superior temporal cortex to the IFG.

The purpose of this study was two-fold: (1) to identify the network of regions associated with complex sentence comprehension in cognitively healthy adults, and (2) to explore how syntactic complexity modulates connectivity within this network. We operationally defined complex sentences as those with arguments in non-canonical order as a result of wh- or NP-movement. Our study used dynamic causal modeling on fMRI data acquired during an auditory sentence-picture verification task to assess the neural mechanisms of processing non-canonical structures with wh- vs. NP-movement. The first hypothesis was that operations involved with processing non-canonical sentences with wh- or NP-movement engage left perisylvian neural networks. Given previous findings of shared syntactic movement processes, we expected to see activity in the left IFG and TPJ. We also predicted that wh- and NP-movement would engage differential activation, reflecting distinct processes, in more focal regions within the left fronto-temporal network. We tested the two aforementioned models of auditory sentence processing which had competing hypotheses for how syntactic movement is processed: via a left fronto-temporal dorsal pathway (Friederici, 2012) or via left temporo-frontal dorsal and ventral pathways (Bornkessel-Schlesewsky and Schlesewsky, 2013).

MATERIALS AND METHODS

Participants

Twenty-one participants (9 females) were recruited from Chicago and surrounding areas to participate in the study and used in the MRI analysis. They were 24–67 years of age ($M = 36.3$; $SD = 13.1$) and had an average of 18.4 years of education ($SD = 2.5$). All participants were right-handed, native speakers of English with normal or corrected-to-normal hearing and vision, and did not have a history of neurological, speech, language, or learning problems. Data from 15 of the 21 participants were used in the connectivity analysis (7 females; age in years: $M = 33.6$, $SD = 10.8$; education in years: $M = 17.9$, $SD = 2.5$) after applying additional exclusionary criteria (see Node Specification section in Effective Connectivity Analysis for details).

All participants passed a MRI safety screening and were compensated for their participation. This study was carried out in accordance with the recommendations of the Human Research Protection Program Plan, Northwestern University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board at Northwestern University.

Procedure

Syntactic processing was assessed using an auditory sentence-picture verification task. Before MRI scanning, participants demonstrated understanding of the task and habituation of the scanning environment with practice inside a mock MRI scanner. At the beginning of each trial, participants saw a visual stimulus followed by an auditory stimulus 500 ms later. They decided with a button press whether a picture matched an auditorily presented sentence using a response box in their left hand, one for their index finger and one for their middle finger. For the *sentence* trials, participants pushed the button under their index finger if the sentence matched the picture, and the button under their middle finger if the sentence and the picture were a mismatch. For low-level auditory-visual processing *baseline* trials, participants were instructed to respond with either button after hearing an auditory stimulus. The response period in sentence trials was longer than that in baseline trials because of differences in task difficulty. Each trial ended in a fixation cross of jittered duration due to the varied length of the auditory stimulus, such that the baseline trials were 6.5 s and sentence trials were 8.5 s. The presentation of events for each trial is detailed in **Figure 1**.

The order of runs for each participant was predetermined using a Latin square design. Since there is evidence that intra-subject reliability is moderate to high across days for language tasks (Wang et al., 2015), some participants completed all four runs on 1day while others completed two runs on day one and two on day two.

Stimuli

Experimental sentences included 16 frequently occurring, transitive verbs that were semantically reversible, had a regular

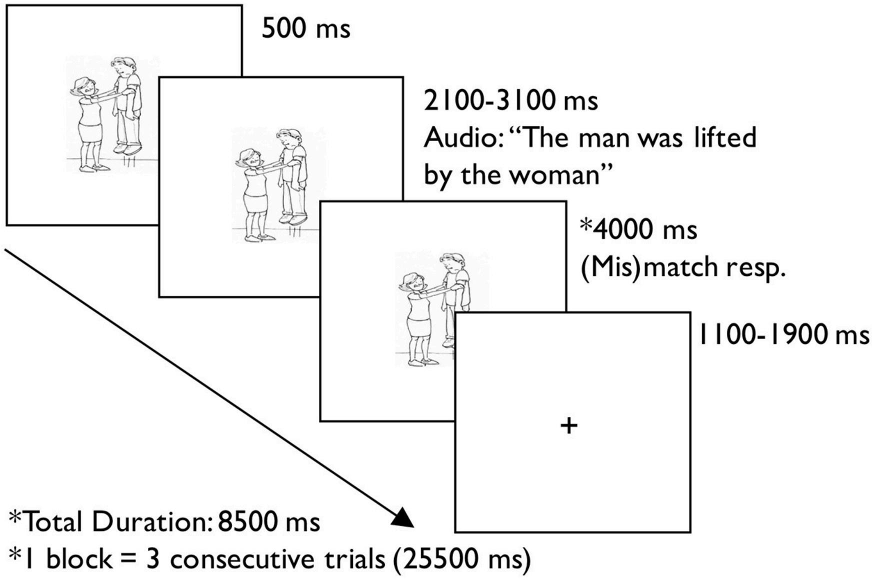


FIGURE 1 | Example trial from the auditory sentence-picture verification task.

passive form (-ed), and were easily picturable. The visual stimuli consisted of 32 black and white drawings, two for each verb. In each pair of drawings, one depicted an agent acting upon the theme, and the other reversed the roles of the participants. The auditory stimuli consisted of 128 unique sentences (see **Appendix**) created by reversing the participant roles in the 16 verbs within four different *sentence* conditions:

1. *Active*: The woman was weighing the boy.
2. *Passive*: The boy was weighed by the woman.
3. *Subject-Cleft (SC)*: It was the woman who weighed the boy.
4. *Object-Cleft (OC)*: It was the boy who the woman weighed.

The first noun phrase in the actives and SCs were agents of the verb and therefore, these sentence types were considered canonical. The first noun phrase in the passives and OCs were not agents of the verb, and therefore, these sentence types were considered non-canonical sentences. The effect of NP-movement can be contrasted by comparing passives and actives as they have a similar syntactic structure. In passives, the moved constituent, *the woman* in (2), occupies an argument position. Similarly, the effect of wh-movement can be contrasted by comparing OCs and SCs because of their similar syntactic structure. Also, both contain the specifier *who* which is linked to agent, *the woman* in (3) and (4). However, in OCs the moved constituent occupies a non-argument position. To control for morphological complexity across conditions, all sentences included past tense verb forms, i.e., past progressive, simple past tense. Given the differences in syntactic structure, cleft sentences were naturally longer than the other two, however, there were no differences in syllable length between the two cleft structures ($M = 9.0$, $SD = 0.85$). A two-sample t-test between active ($M = 7.91$, $SD = 0.87$), and passive sentences ($M = 8.0$, $SD = 0.84$) also confirmed no differences in syllable length for these structures [$t_{(93.9)} = -0.71$, *n.s.*]. Auditory recordings were made by a female speaker with typical prosody, normalized to a consistent sound level, and adjusted to a slightly slower than normal speech rate ($M = 3.36$ syllables per second, $SD = 0.30$) using the program Audacity® version 2.0.0[1].

A low-level auditory-visual processing *baseline* condition was also included. For this condition, 16 time-reversed audio files (four randomly selected from each sentence condition) and 16 baseline visual stimuli were created. Eight of the baseline visual stimuli came from eight randomly selected pictures that were partitioned into 8 × 8 grids and scrambled. The other eight baseline visual stimuli were 180° rotated versions of the eight scrambled pictures.

The experiment included 288 trials: 192 sentences (48 per type) and 96 baseline. Trials were blocked by condition such that there were 3 trials per block, which yielded 64 sentence blocks, and 32 baseline blocks. The trials were pseudorandomized such that the same verb was presented at least one block apart, half of the trials for each sentence condition matched the presented picture and the other half mismatched the picture, and all agents in each sentence block were not all the same gender. The blocks were grouped into four runs, such that runs A and B consisted of *passive*, *active*, and *baseline* conditions, and runs C and D consisted of *OC*, *SC*, and *baseline* conditions.

MRI Data Acquisition

MRI data were acquired using a 32-channel head coil on a 3 Tesla Siemens TRIO system. To obtain an anatomical image of the brain, T1-weighted three-dimensional multi-planar rapid acquisition gradient echo (MPRAGE) sequences recorded 176

---

[1] Audacity® software is copyright © 1999–2014 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

slices with a voxel size of 1.0 × 1.0 × 1.0 mm, using a repetition time of 2,300 ms, echo time of 2.91 ms, a flip angle of 9°, and field of view of 256 mm.

During the experimental task, functional MRI blood oxygen level dependent (BOLD) data were acquired such that each image consisted of 41 slices and a voxel size of 1.7 × 1.7 × 3.0 mm. These images were recorded using a repetition time of 2,400 ms, echo time of 20 ms, a flip angle of 90° and a field of view of 220 mm resulting in a matrix size of 129 × 129. All imaging was conducted at the Department of Radiology's Center for Translational Imaging at Northwestern University.

Behavioral Data Analysis

Accuracy and reaction time (RT) data from the auditory-sentence picture verification task were analyzed using the R software version 3.5.1. With respect to assumptions of normality, non-parametric sign tests of the accuracy data and parametric paired t-tests of the RT data from participants were conducted to elucidate any differences between sentence conditions: active vs. passive, OC vs. SC, and canonical vs. non-canonical.

FMRI Data Analysis

Individual analysis of anatomical and functional neuroimaging data was conducted on the Northwestern University Neuroimaging Data Archive, which allowed for automatic and optimized preprocessing and first-level statistical data analysis pipeline. Anatomical images went through the following preprocessing steps: skull-stripping, segmentation, registration, and normalization. Preprocessing for functional scans began with despiking, censoring data if framewise displacement was 0.5 mm or greater, slice-time correction, co-registration of the anatomical scan to the mean functional volume, regressing signal from white matter and cerebrospinal fluid, normalization of the anatomical and functional scans using the VBM/DARTEL template in Montreal Neurological Institute (MNI) space (2 × 2 × 2 mm resolution), and smoothing using a 6 mm Gaussian kernel. Scripts from Analysis of Functional NeuroImages (AFNI), FMRIB Software Library (FSL), and Statistical Parametric Mapping 8 (SPM8) were utilized for preprocessing. More details about this implementation can be found in Alpert et al. (2016).

In the first-level statistical analysis, a high pass filter of 128 s was used to eliminate scanner drift. A general linear model containing passive, active, OC, SC, and baseline conditions was specified and estimated in SPM8. Activation for general sentence processing was found with the contrast (*Passive* + *Active* + *OC* + *SC*) > *Baseline*, alternatively referred to as *Sentences* > *Baseline*. Activation for non-canonical compared to canonical sentence processing was identified with the contrast (*Passive* + *OC*) > (*Active* + *SC*), alternatively referred to as *Noncanonical* > *Canonical*. Preferential activation for processing wh-movement was defined by the (*OC* > *SC)* > (*Passive* > *Active*) contrast, or *Wh* > *NP-movement*, whereas (*Passive* > *Active*) > (*OC* > *SC*) was used for isolating activation associated with NP-movement, or *NP* > *Wh-movement*.

At the group level, regions involved with all sentences (both canonical and non-canonical), non-canonical sentence processing, and wh- vs. NP-movement were identified. For each contrast of interest, a one-sample t-test of the group's images was conducted with age as a covariate in SPM8. This statistical analysis yielded a binary mask of active voxels common to all participants and an image of the voxel-wise variance of error from the t-test. The binary mask and variance of error image were input into 3dFWHMx program from Analysis of Functional NeuroImage (AFNI) to estimate noise smoothness in x-, y-, and z-directions. This was done by fitting the data to a Gaussian plus mono-exponential mixed model because functional MRI data do not have a Gaussian-shaped autocorrelation function, as previously assumed (Eklund et al., 2016). To determine the maximum size of false positive (noise-only) clusters, the estimated noise smoothness, binary mask, and image of variance of error were used in AFNI's 3dClustSim to calculate cluster-defining thresholds at $\alpha = 0.01$ level of significance and a specified voxel-wise threshold ($p < 0.001$, uncorrected). This program used Monte-Carlo simulations given a specified voxel-wise level of significance. The cluster-defining threshold was reported using 2-sided thresholding since we were interested in both directions of the contrasts of interest, and with first-nearest neighbor clustering, because it produces the most conservative result. The Harvard-Oxford atlas was used to throughout the study to label peak activation.

Effective Connectivity Analysis

Effective connectivity describes how neural activity from one region influences neural activity of another region. Dynamic Causal Modeling (DCM) is a hypothesis-driven method for estimating effective connectivity using functional MRI data (Friston et al., 2003). DCM is particularly useful for testing hypotheses about the influences of particular connections within a neural network, e.g., how sensory stimuli or experimental tasks modulate neuronal interactions. The method is dynamic because it uses differential equations to estimate connectivity and task-induced neuronal interactions, and causal because directionality can be specified (Seghier et al., 2010; Stephan et al., 2010). For task-based functional MRI studies, experimental conditions serve as input into the model by either driving neural activity throughout the network and/or modulating connectivity between regions.

The first step in DCM is selecting the nodes in the network and identifying connections. For this study, node selection and connectivity were guided by neurocognitive models of sentence comprehension. The DCM12 toolbox in SPM12 was used for effective connectivity analysis. Each model was specified using binary values in three matrices: the A-matrix, B-matrix, and C-matrix. The A-matrix is an n-by-n square matrix representing the intrinsic connectivity (i.e., in the absence of external input) between the n nodes. The B-matrix is an n-by-n-by-c matrix representing how c experimental conditions causes a change in the rate of neural activity between the n nodes. The C-matrix is an n-by-i matrix representing the i external inputs that would affect the rate of change of neural activity of n nodes which consequently drives activity within the model, e.g., the "driving input." This matrix triplet, signifying one model, and the fMRI data were inputs to the DCM12 toolbox to estimate model parameters and calculate model fit using Bayesian statistics.

Node Selection

For the present study, nodes in the network were specified from group peaks identified in the *Noncanonical* > *Canonical* contrast ($p < 0.001$, uncorrected; $k > 25$) masked by the *All Sentences* > *Baseline* contrast ($p < 0.001$, uncorrected) to isolate sentence processing regions involved in processing both wh- and NP-movement. One of the peak coordinates (−48, 22, 22) did not have a label in the Harvard-Oxford Atlas, but it was labeled as left pars opercularis of the inferior frontal gyrus (LIFGop) because the cluster extended primarily into that region. This resulted in peaks within the LIFGop, left medial superior frontal gyrus (LSFGm), and left middle frontal gyrus (LMFG), and left posterior superior temporal gyrus (LSTGp). The same contrast and mask were applied to all first-level analyses to identify suprathreshold voxels ($p < 0.05$, uncorrected) within a 12 mm radius sphere centered at each of the group peaks (or local maxima if <10 suprathreshold voxels were yielded using the group peak). Subject-specific eigenvariates were adjusted for effects-of-interest and extracted from a modified general linear model using these suprathreshold voxels as a mask. Regions were excluded from the DCM analysis if consistent activation was not observed across subjects. In addition, subjects were excluded from the DCM analysis if suprathreshold voxel-wise activation was <10 voxels for at least one of the resulting nodes. These exclusionary criteria were imposed to decrease the likelihood of incorporating noisy data during model estimation. The modified general linear model concatenated all 4 runs, modeled the 5 conditions, and regressed for all 4 runs and linear drift for each run.

Model Specification and Estimation

In the present study, neuronal connections within models were assumed to be bilinear and deterministic (see Seghier et al., 2010 for a description of all model specification options) which are appropriate for neurologically normal participants. Further, two-state neuronal equations were used to improve model estimation by quantifying the interaction between inhibitory and excitatory neuronal subpopulations within a given region (Marreiros et al., 2008). Unlike modeling with one-state neuronal equations, positive constraints (or priors) for between-region connections and negative constraints on within-region connections were implemented for determining intrinsic connectivity (e.g., A-Matrix). Two-state DCM also estimated the proportional increase or decrease from intrinsic connectivity between regions caused by task-induced perturbations (e.g., B-Matrix). Parameters for intrinsic connectivity and modulations were log scaled. For statistical analysis, they were exponentially transformed such that a value of 1 represents no neural rate of change from region X to region Y; a value <1 represents a decrease in neural rate of change from region X to region Y; and values >1 represent an increase in the neural rate of change from region X to region Y. Parameters for external driving input (e.g., C-Matrix) are estimated in hertz.

Two sets of models for each movement type were specified and estimated, such that *All Sentences* > *Baseline* contrast was indicated as the driving input (i.e., sentence processing driving neural activity within the network) and either the *OC* > *SC* contrast (i.e., processing wh-movement) or the *Passive* > *Active* contrast (i.e., processing NP-movement) modulated connectivity between regions. Specifying all three contrasts within the same GLM would leave no variance for the model. Therefore, it was not possible to directly compare models of wh- and NP-movement in a statistical way. **Figure 2** illustrates the bidirectional intrinsic connections specified between the LIFGop and LSTGp, as they are connected by the superior longitudinal fasciculus, and between the LIFGop and LSFGm, by way of the frontal aslant tract (Catani et al., 2013; Dick et al., 2014; Martino and Lucas, 2014).

To create different plausible models, driving input either entered the LIFGop, LSTGp, or both the LIFGop and LSTGp. In addition, syntactic movement either modulated the connection from the LIFGop to the LSTGp (LIFGop-LSTGp), LSTGp-LIFGop, or both connections. Syntactic movement was also modeled such that it either did not modulate connectivity between the LIFGop and LSFGm, it only modulated LIFGop-LSFGm, or it modulated both LIFGop-LSFGm and LSFGm-LIFGop. Models were excluded if connectivity to, but not from, the driving input region was modulated by syntactic movement, as this overemphasizes the neural activity in the driving input region. This resulted in two sets of 21 models (see **Table 1** for details). Each set contained models that estimated the effect of wh-movement or NP-movement on connectivity. Model evidence, the probability of observing the fMRI data given the model's specifications, was calculated for every single model.

Bayesian Model Selection and Averaging

For each type of syntactic movement, the 21 models were grouped into three different families: those with driving input into the LIFGop, into the LSTGp, or into both. A random-effects family-wise Bayesian Model Selection (BMS) was performed and the winning family was that which had the highest exceedance

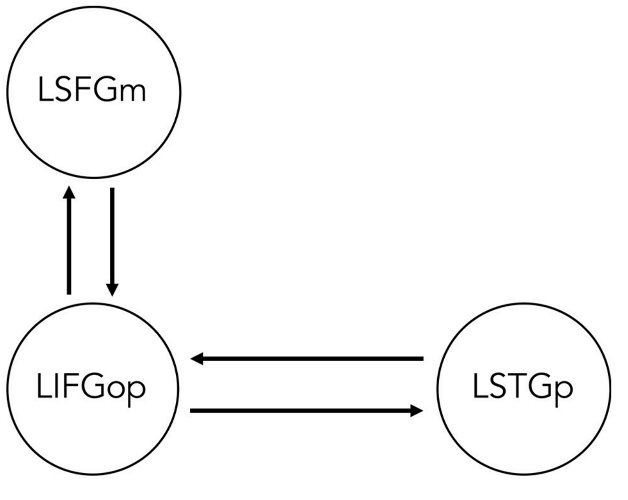


FIGURE 2 | Base dynamic causal model of network of regions involved in non-canonical sentence processing for sentence-picture verification task. LSFGm, Left superior frontal gyrus, medial part. LIFGop, left inferior frontal gyrus, opercular part. LSTGp, Left superior temporal gyrus, posterior part.

TABLE 1 | Full model space for DCM analysis (1 = included, 0 = not included).

	Driving input		Modulated by syntactic movement			
Model ID	**IFGop**	**STGp**	**IFG-STG**	**STG-IFG**	**IFG-mSFG**	**mSFG-IFG**
M1	1	1	1	1	1	1
M2	1	1	1	1	1	0
M3	1	1	1	1	0	0
M4	1	1	1	0	1	1
M5	1	1	0	1	1	1
M6	1	1	1	0	1	0
M7	1	1	0	1	1	0
M8	1	1	1	0	0	0
M9	1	1	0	1	0	0
M10	1	0	1	1	1	1
M11	1	0	1	1	1	0
M12	1	0	1	1	0	0
M13	1	0	1	0	1	1
M14	1	0	1	0	1	0
M15	1	0	1	0	0	0
M16	0	1	1	1	1	1
M17	0	1	1	1	1	0
M18	0	1	1	1	0	0
M19	0	1	0	1	1	1
M20	0	1	0	1	1	0
M21	0	1	0	1	0	0

probability. The exceedance probability is the likelihood that a particular family of models, compared to the other families, generated the data of a randomly selected participant from the group. Therefore, the sum of the exceedance probabilities of all families in the BMS equaled 1. The winning family was subject to Bayesian Model Averaging (BMA) to obtain a model average which contained parameters weighted by the posterior probability of each contributing model within the family. Subject-specific parameters from the A-, B-, and C-matrices were entered into one-sample t-tests and corrected for multiple comparisons via false discovery rate (FDR) to determine whether the estimated intrinsic connectivity was significantly different from the prior constraint, whether the estimated modulatory effect was significantly different from no effect, and whether the estimated effect of the driving input was significantly >0 Hz. To determine whether particular connections were significantly stronger than others, all intrinsic connections and modulated connections were entered into separate general linear models in order to conduct simultaneous pair-wise comparisons.

RESULTS

Behavioral Results

A significant canonicity effect in accuracy was found such that participants were more accurate across the two canonical sentence types ($median = 1$) vs. the two non-canonical sentence types ($median$=0.98), $p < 0.005$. This effect was primarily driven by higher accuracy for SC sentences ($median = 0.99$) compared to OC sentences ($median = 0.98$), $p = 0.06$. No statistically significant difference in accuracy was observed between passives and actives. With the exception of one participant whose accuracy ranged from 83 to 94%, all other participants were 90% accurate or greater across all 4 structures. Participants were also significantly quicker to respond [$t_{(20)} = 8.55$, $p < 0.005$] to both canonical sentences types ($M = 2776$ ms, SD =2 45) compared to the two non-canonical sentences types (M = 2,934, SD = 282). They had a significantly faster RT for actives (M = 2,642, SD = 247) compared to passives [M = 2,759, SD = 289; $t_{(20)} = 5.36$, $p < 0.005$] and for SCs (M = 2,903, SD = 260) compared to OCs [M = 3,092, SD = 315; $t_{(20)} = 5.76$, $p < 0.005$]. Direct comparison of the non-canonical structures showed no difference in accuracy [$t_{(20)} = 0.37$, $p = 0.71$], but a significantly faster reaction time for passives compared to OCs [$t_{(20)} = 5.91$, $p < 0.001$].

FMRI Results

Task performance elicited large clusters of activation primarily in the left hemisphere for general sentence processing (*All Sentences > Baseline*) (cluster-defining threshold was $k = 61$; **Figure 3**, top row, red-yellow gradient). Peak activations were in the left pars triangularis of the IFG, middle frontal gyrus (MFG), supplementary motor area (SMA), temporal pole, SPL, superior LOC (LOCs), occipital pole, cerebral white matter, and right inferior lateral occipital cortex (LOCi) (**Table 2**). The Harvard-Oxford atlas did not have a label for the peak located at (10, −68, −24), but it appeared to be located within the right cerebellum. The opposite contrast, activation for the baseline condition compared to the all sentence conditions, yielded peak activation in the bilateral paracingulate gyrus, SFG, right planum temporale, lingual gyrus, posterior supramarginal gyrus (SMGp), MTGp, left planum polare, and posterior cingulate gyrus (**Figure 3**, top row, blue-green gradient). No significant regions of activation were elicited for canonical compared to non-canonical sentences and for NP- compared to wh-movement.

Non-canonical sentence processing elicited peak activity (cluster-defining threshold was $k = 43.4$; **Figure 3**, middle row, red-yellow gradient) in the left hemisphere. This included the left pars opercularis of the IFG, MFG, paracingulate gyrus, MTGp, LOCs, and occipital fusiform gyrus (**Table 2**).

Contrasting wh- compared to NP-movement structures (cluster-defining threshold was $k = 42.1$; **Figure 3**, bottom row, red-yellow gradient) also showed a left hemisphere network of perisylvian regions. Wh-movement elicited peak activity in the medial SFG (SFGm), insular cortex, pars opercularis of the IFG, MTGp, and LOCs (**Table 2**).

Effective Connectivity Results

Data from 15 of the 21 participants were used in the DCM analysis (7 females; age in years: M = 33.6, SD = 10.8; education in years: M = 17.9, SD = 2.5). Six subjects were excluded because they had <10 suprathreshold voxels for at least one of three resulting nodes. **Figure 4**. illustrates the *Noncanonical > Canonical* contrast ($p < 0.001$, uncorrected; $k > 25$) masked by the *Sentences > Baseline* contrast ($p < 0.001$, uncorrected). **Table 3** reports the peak activation in MNI

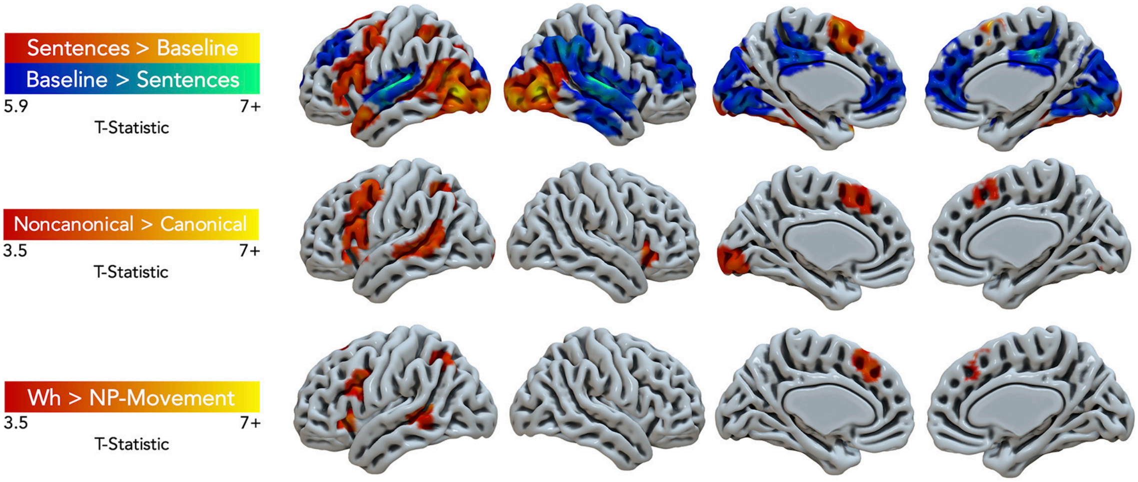


FIGURE 3 | Significant fMRI activation (uncorrected voxelwise $p < 0.001$) of the contrasts *Sentences > Baseline* and vice-versa (corrected cluster-defining threshold $k > 61$), *Non-canonical > Canonical* ($k > 43.4$), and *Wh > NP-movement* ($k > 42.1$) from healthy adult participants.

space which included peaks within the LIFGop, left posterior superior temporal gyrus (LSTGp), left medial superior frontal gyrus (LSFGm), and left middle frontal gyrus (LMFG). First-level analyses revealed inconsistent activation within the MFG across subjects demonstrating that this region's neural response was driven by a subset of participants. Therefore, the MFG was not included as a node in the DCM analysis.

Wh-Movement Models

An initial random-effects BMS was conducted among the 21 models and there was no clear winning model (highest exceedance probability = 0.38, next highest exceedance probability = 0.20). Provided these results, a random-effects family-wise BMS was conducted and is illustrated in **Figure 5** *(top panel)*. Among the 3 model families, the winning family was the set of models with driving input into the LIFGop (exceedance probability = 0.73) with the next best winning family being the set of models with driving input into the LSTGp (exceedance probability = 0.25).

BMA was conducted across the 6 models with input into the LIFGop (models 10–15, see **Table 1**) yielding averaged model parameters weighted by their posterior probability. Inspection of individual data resulted in exclusion of one participant because their estimated parameters were >3 standard deviations from the mean. **Figure 5** (bottom left panel) displays the parameters for intrinsic connections in which the red-dotted line, equal to the value of 1, denotes no estimated difference from the prior. Estimated parameters for intrinsic connections were greater than the prior for the LIFGop-LSFGm ($M = 1.24$, $SD = 0.33$; $p < 0.05$, uncorrected), LIFGop-STGp ($M = 1.27$, $SD = 0.40$; $p < 0.05$, uncorrected), LSFGm-LIFGop ($M = 1.06$, $SD = 0.10$; $p < 0.05$, uncorrected), and LSTGp-LIFGop ($M = 1.13$, $SD = 0.25$; *n.s.*). Simultaneous pairwise comparisons of intrinsic connections revealed a trend toward significance ($p < 0.08$, uncorrected) between the LIFGop-STGp and LSFGm-LIFGop [$t_{(3,\,52)} = -1.87$]. Driving input into the LIFGop was significantly >0 ($M = 0.03$ Hz, $SD = 0.02$, $p < 0.001$).

Figure 5 (bottom right panel) displays the parameters for connections modulated by wh-movement in which the red-dotted line, equal to the value of 1, denotes no estimated difference from the intrinsic connection when processing wh-movement from region X to region Y. Wh-movement significantly modulated all connections after a FDR correction for multiple comparisons, $p < 0.05$: LIFGop-LSFGm ($M = 1.24$, $SD = 0.27$), LIFGop-LSTGp ($M = 1.22$, $SD = 0.38$), LSFGm-LIFGop ($M = 1.04$, $SD = 0.05$), and LSTGp-LIFGop ($M = 1.21$, $SD = 0.28$). Simultaneous pairwise comparisons of modulatory connections also revealed a trend toward significance ($p < 0.08$, uncorrected) between the LIFGop-LSFGm and LSFGm-LIFGop, $t_{(3,\,52)} = -2.00$, and between the LIFGop-STGp and LSFGm-LIFGop, $t_{(3,\,52)} = -1.79$. **Table 4** summarizes the statistical analyses of the mean parameter estimates from the BMA models modulated by wh-movement in which a value of 1 denotes no estimated perturbation in neural rate of change intrinsically or in response to processing wh-movement from one region to the other.

NP-Movement Models

An initial random-effects BMS was conducted among the 21 models and there was also no clear winning model for the NP-movement models (highest exceedance probability = 0.46; next highest exceedance probability = 0.29). Provided these results, a random-effects family-wise BMS was conducted and is illustrated in **Figure 6** (top panel). Among the 3 model families, the winning family was the set of models with driving input into the LIFGop (exceedance probability = 0.82) with the next best winning family being the set of models with driving input into the LSTGp (exceedance probability = 0.18).

Similar to the wh-movement results, BMA was conducted across the 6 models with input into the LIFGop for

TABLE 2 | Summary of peak activation in MNI space.

Contrast	L/R	Peak location	k	T	x	y	z
Sentences > baseline	L	IFGtri	480	6.2	−50	28	18
	R	SMA	217	9.0	−4	8	60
	L	Temporal pole	127	7.7	−54	8	−18
	L	Cerebral white matter	112	6.3	−14	−2	14
	L	MFG	432	6.7	−30	−4	56
	L	Superior parietal lobule (SPL)	190	7.3	−28	−46	44
	R	N/A	104	9.2	10	−68	−24
	L	LOCs	73	6.5	−26	−70	32
	R	LOCi	2,285	11.2	48	−76	8
	L	Occipital pole	3,271	11.3	−26	−100	−8
Baseline > sentences	R	Paracingulate gyrus	462	−6.9	6	50	4
	R	Paracingulate gyrus	131	−6.3	4	42	28
	L	SFG	429	−6.9	−20	28	40
	R	SFG	1,844	−8.8	20	26	56
	L	Planum polare	1,910	−13.9	−48	−10	−2
	R	Planum temporale	2,347	−10.8	60	−12	6
	R	MTGp	146	−6.9	58	−12	−30
	R	Posterior cingulate gyrus (pCG)	730	−9.3	2	−28	32
	R	SMGp	941	−10.3	56	−40	40
	R	Lingual gyrus	2,069	−9.7	18	−74	−4
Non-canon > canonical	R	Frontal orbital cortex	63	5.1	38	22	−2
	L	IFGop	271	6.8	−48	22	22
	L	MFG	132	6.2	−50	16	38
	L	Paracingulate gyrus	67	5.4	−2	14	52
	L	MTGp	308	10.5	−52	−36	0
	L	LOCs	50	6.0	−34	−62	46
	R	N/A	44	6.0	10	−74	−24
	L	Occipital fusiform gyrus (OFG)	120	5.6	−14	−86	−12
Canonical > Non-canon		No significant clusters of activation					
Wh > NP-Movement	L	SFGm	47	5.7	−6	30	46
	L	Insular cortex	70	8.7	−32	24	−2
	L	IFGop	167	7.6	−36	16	24
	L	MTGp	49	5.3	−56	−42	2
	L	LOCs	68	6.3	−36	−58	46
NP > Wh-Movement		No significant clusters of activation					

NP-movement (models 10–15, see **Table 1**). **Figure 6** (bottom left panel) displays the parameters for intrinsic connections in which the red-dotted line, equal to the value of 1, denotes no estimated difference from the prior. Parameters for intrinsic connections were greater than the prior for all connections, $p(\text{FDR}) < 0.05$: LIFGop-LSFGm ($M = 1.27$, $SD = 0.31$), LIFGop-LSTGp ($M = 1.27$, $SD = 0.32$), LSFGm-LIFGop ($M = 1.06$, $SD = 0.07$), and LSTGp-LIFGop ($M = 1.14$, $SD = 0.21$). Simultaneous pairwise comparisons of intrinsic connections revealed a difference ($p < 0.05$, uncorrected) between the LIFGop-LSFGm and LSFGm-LIFGop, $t_{(3,56)} = -2.38$, and between the LIFGop-LSTGp and LSFGm-LIFGop, $t_{(3,56)} = -2.34$. Driving input into the LIFGop was significantly >0 ($M = 0.03$ Hz, $SD = 0.02$; $p < 0.001$). **Figure 6** (bottom right panel) displays the parameters for connections modulated by NP-movement in which the red-dotted line, equal to the value of 1, denotes no estimated difference from the intrinsic connection when processing NP-movement from region X to region Y. NP-movement modulated LIFGop-LSFGm ($M = 1.49$, $SD = 0.75$; $p(\text{FDR}) < 0.05$), LIFGop-STGp ($M = 1.36$, $SD = 0.50$; $p(\text{FDR}) < 0.05$), LSFGm-LIFGop ($M = 1.07$, $SD = 0.10$; $p(\text{FDR}) < 0.05$), and LSTGp-LIFGop ($M = 1.21$, $SD = 0.28$; $p < 0.05$, uncorrected). Simultaneous pairwise comparisons revealed a difference between the LIFGop-LSFGm and LSFGm-LIFGop ($t_{(3,56)} = -1.91$; $p = 0.06$, uncorrected). **Table 5** summarizes the statistical analyses of the mean parameter estimates from the BMA models modulated by NP-movement in which a value of 1 denotes no estimated perturbation in neural rate of change intrinsically or in response to processing NP-movement from one region to the other.

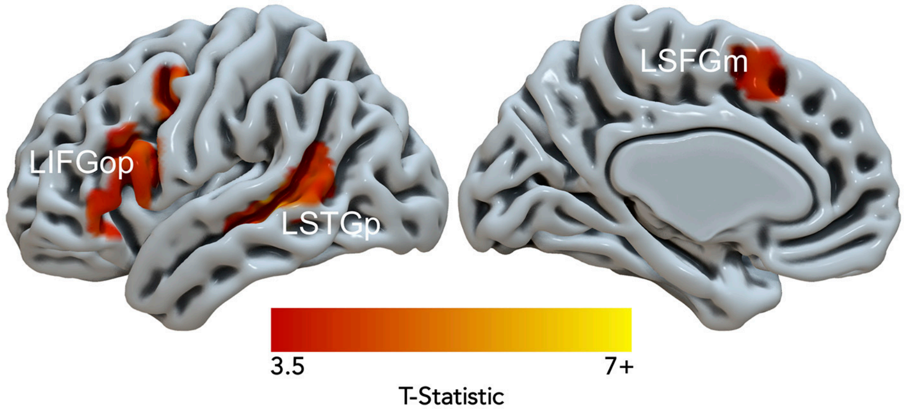


FIGURE 4 | Significant fMRI activation (uncorrected voxelwise $p < 0.001$) of *Non-canonical* > *Canonical* (corrected cluster-defining threshold $k > 61$) masked by *Sentences* > *Baseline* ($k > 61$) from healthy adult participants.

TABLE 3 | Peak and sub-peak activation for Non-canonical > canonical masked by sentences > baseline.

L/R	Location of peaks and submaxima peaks (indented)	k	T	x	y	z
L	Inferior frontal gyrus, opercular part (LIFGop)	121	6.78	−48	22	22
	- Inferior frontal gyrus, opercular part		5.24	−54	16	18
	- Inferior frontal gyrus, triangular part		5.04	−52	20	−2
L	(Medial) Superior frontal gyrus (LSFGm)	28	5.19	−6	16	50
	- (Medial) Superior frontal gyrus		4.46	−4	10	58
L	Middle frontal gyrus (LMFG)	33	5.10	−42	2	48
L	Superior temporal gyrus, posterior division (LSTGp)	148	10.48	−52	−36	0
	- Angular gyrus		5.54	−54	−52	10
	- Middle temporal gyrus, temporo–occipital part		5.24	−48	−48	6

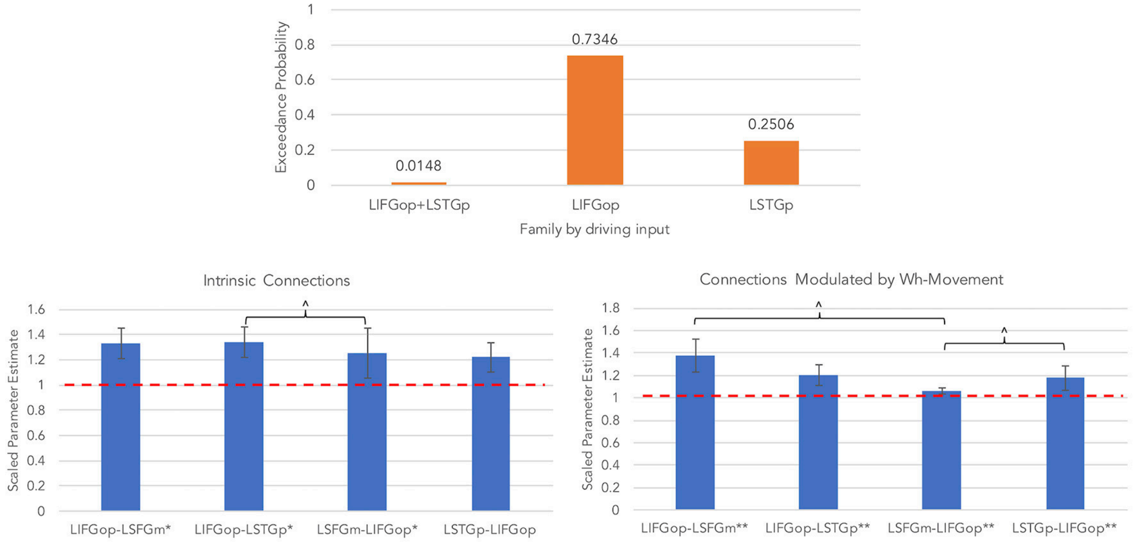


FIGURE 5 | Results for wh-movement models (**$p < 0.05$, FDR-corrected; *$p < 0.05$, uncorrected; ^$p < 0.08$, uncorrected).

DISCUSSION

The primary aims of this study were to identify the neural network associated with the comprehension of complex sentences and to explore how syntactic complexity modulated connectivity within this network. Using an auditory sentence-picture verification fMRI task, this study demonstrated that non-canonical sentences with wh-movement elicit greater neural activity than those with NP-movement, though both types of movement modulate neural connectivity in a similar manner. While findings from the fMRI analysis support the idea that processing the wh-movement operation requires more neurocognitive resources than the NP-movement, results from the connectivity study suggest that both movement operations may undergo the same stages of processing.

First, results from the fMRI analysis revealed that all sentence conditions compared to the baseline condition (*All Sentences* > *Baseline*) yielded a mostly left hemisphere network with peaks in the left pars triangularis of the IFG, right SMA, left temporal pole, MFG, SPL, LOCs, right LOCi, and left occipital pole. As many of these regions have been previously found in neuroimaging studies of sentence processing (see, Friederici, 2011 for a comprehension review), these results are validating. Bilateral occipital activation was observed in visual association cortex which responds more to complex visual representations than simple visual stimuli (Van Essen and Maunsell, 1983). The opposite contrast (*Baseline* > *All Sentences*) resulted in a bilateral network of regions with peaks in bilateral paracingulate gyrus, SFG, left planum polare, right planum temporale, MTGp, posterior cingulate gyrus, SMGp, and lingual gyrus. Activation in bilateral STG fell within the primary auditory cortex and likely reflected the contrast between hearing reversed speech vs. spoken sentences (Skipper, 2014). The other medial and right hemisphere regions were previously reported when comparing less to more cognitively demanding tasks (Raichle and Snyder, 2007). Therefore, they may reflect differences in cognitive functioning or effort between *baseline* and *sentence* conditions.

In line with previous reports of noncanonical sentence processing (Bornkessel et al., 2005; Caplan et al., 2008; Thompson et al., 2010b; Bornkessel-Schlesewsky et al., 2012; Makuuchi

TABLE 4 | Mean (and standard deviation) of subject-specific scaled BMA parameters for wh-movement models.

Connection	Intrinsic connectivity	Modulation by Wh-Mov
IFGop-SFGm	*1.24 (0.33)	**1.23 (0.27)
IFGop-STGp	*1.27 (0.40)	**1.22 (0.38)
SFGm-IFGop	*1.06 (0.10)	**1.04 (0.05)
STGp-IFGop	1.13 (0.25)	**1.21 (0.08)

***p < 0.05, FDR-corrected; *p < 0.05, uncorrected.*

TABLE 5 | Mean (and standard deviation) of subject-specific BMA parameters for NP-movement models.

Connection	Intrinsic connectivity	Modulation by NP-Mov
IFGop-SFGm	**1.27 (0.31)	**1.49 (0.75)
IFGop-STGp	**1.27 (0.32)	**1.36 (0.50)
SFGm-IFGop	**1.06 (0.07)	**1.07 (0.10)
STGp-IFGop	**1.14 (0.21)	^1.44 (0.80)

***p < 0.05, FDR-corrected; *p < 0.05, uncorrected; ^p < 0.08, uncorrected.*

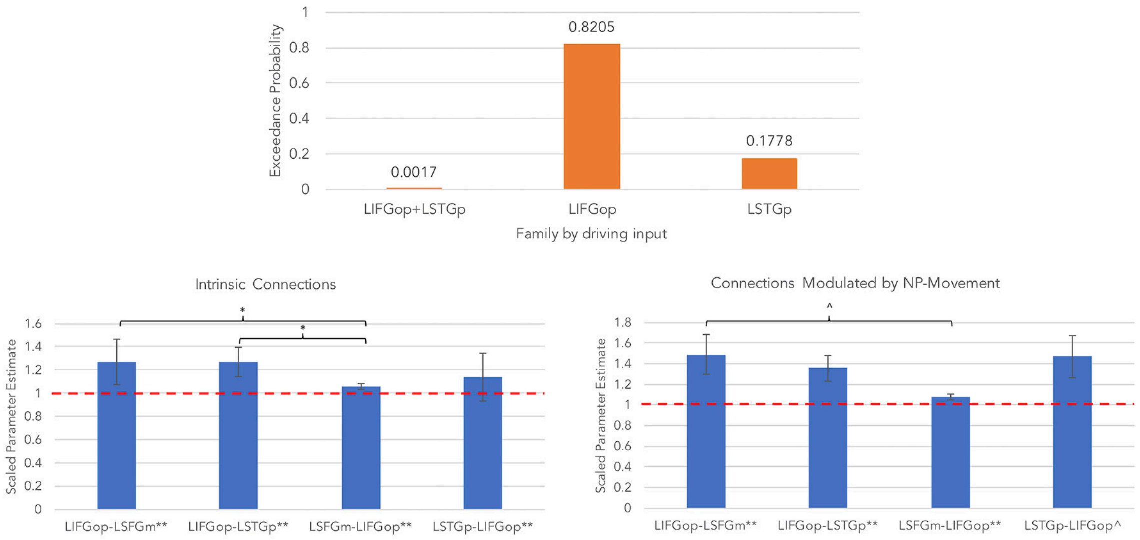


FIGURE 6 | Results for NP-movement models (**$p < 0.05$, FDR-corrected; *$p < 0.05$, uncorrected; ^$p < 0.08$, uncorrected).

et al., 2012; Mack et al., 2013), this experiment yielded a mostly left hemisphere network with peaks in pars opercularis of the IFG, MFG, paracingulate gyrus, MTGp, and LOCs during non-canonical sentence comprehension when compared to canonical sentences (*Noncanonical* > *Canonical*), while no significant activation was found for the opposite contrast (*Canonical* > *Noncanonical*). Non-canonical sentence processing, examined by combining the wh- and NP-movement contrasts (e.g., *OC* + *Passive* > *SC* + *Active*), requires forming a dependency between the moved constituent and the trace site resulting in reactivation of the filler after the verb is encountered to initiate re-assignment of its thematic role. We found that the left IFG and MTGp were two regions active for this contrast, which is consistent with both models of auditory sentence comprehension tested (Friederici, 2012; Bornkessel-Schlesewsky and Schlesewsky, 2013). Both predict involvement of these regions in processing non-canonical sentences, but disagree with regard to their function. According to the model by Bornkessel-Schlesewsky and Schlesewsky (2013), lexical-semantic processing first takes place in left posterior temporal regions followed by combinatorial syntactic and thematic processing in dorsal and ventral pathways, respectively, to the left IFG where these two types of information are integrated. On the other hand, Friederici (2012) claims that the left IFG is involved with assigning grammatical relations between syntactic constituents, which precedes involvement of the left posterior superior temporal cortex in re-assigning thematic roles.

Our results also showed that wh-movement elicits greater activity in left inferior frontal and posterior temporal cortices compared to NP-movement which may reflect greater processing resources for thematic role assignment in the context of the double-dependency seen in wh-movement. This novel finding (*Wh* > *NP-movement*) revealed a left perisylvian network with peak activity in the SFGm, insular cortex, pars opercularis of the IFG, MTGp, and LOCs, but no significant activity vice-versa, which provides support for both representational and processing accounts of wh-movement. Representationally, object-cleft sentences entail movement across clausal boundaries, i.e., a type of A-bar movement in which the moved constituent, *who*, occupies a non-argument position, the specifier position of the Complement Phrase. This results not only in a co-referential relation between the moved constituent (*who*) and the trace (as in NP-movement structures), but also between *who* and the head noun of the matrix clause. In contrast, NP-movement is a type of A-movement and it occurs when the displaced constituent, the filler, occupies an argument position and leaves behind a trace. In the passive sentences used in the study, the filler occupies the subject position in the syntactic frame because it is an argument of the verb. The trace forms a direct dependent (co-referential) relationship with the noun phrase. Although both object-cleft sentences and passives were highly accurate, reaction times were longer for the wh-movement structures compared to the NP-structures which provides additional evidence for processing differences. Wh-movement elicited activity in the left pars opercularis of the IFG and the left insular cortex, consistent with the model of auditory sentence comprehension by Friederici (2012) describing the left pars opercularis' role in processing higher-order syntactic relations.

In addition, *Wh* > *NP-movement* elicited activity in a subset of the regions observed in the *Noncanonical* > *Canonical* contrast, namely the SFGm and LOCs, which may reflect processes shared between the two movement types, but require additional computational resources for wh-movement. From a linguistic standpoint, the left frontal activation has been reported for effortful sentence comprehension (Adank, 2012a,b) as well as word sequencing (Crozier et al., 1999; Alario et al., 2006) which is more relevant for OC sentences due to the non-canonical word order. The peak within the LOCs is rostrally adjacent to both the SPL and the angular gyrus. Activation in this area, particularly the angular gyrus, has been associated with processing thematic relations between words (Kalénine et al., 2009; Boylan et al., 2015, 2017; Lewis et al., 2015). Thompson and Meltzer-Asscher (2014) argues that the function of the AG is the retrieval or argument structure information within their model of verb argument structure processing. However, some claim that these regions instead play a domain-general role in language processing because they fall outside of the more conventional left frontotemporal syntax processing network (see Campbell and Tyler (2018) for more details on this argument). The left frontal activation in this study is located dorsally to the inferior frontal gyrus, putatively involved in syntactic processing, and has been associated with the domain-general multiple demand network (Campbell et al., 2016). In addition, the left LOCs and the area rostrally adjacent have been previously linked to attentional processing (Dreher and Grafman, 2003; Mizuno et al., 2012). Further investigation is required to determine whether these regions are specifically relevant for syntactic processing or support language processing in general.

The effective connectivity analysis demonstrated that syntactic movement modulated both temporofrontal and frontotemporal pathways, where external input to the LIFGop drove neural activity throughout the network. Two related studies previously examined how syntactic complexity modulated connectivity between language regions. den Ouden et al. (2012) used a similar auditory sentence-picture verification task to examine how syntactic complexity modulated the network. Their result was a model in which the LIFG's response to sentences drove network activity, and OC sentences modulated the connection from the LIFG to the LSTGp. Similarly, Makuuchi and Friederici (2013) employed a sentence verification task to determine how complex sentences modulated activity in the reading network. Because this was a reading task, the driving input was to the left fusiform gyrus. Syntactic complexity, however, modulated activity from the inferior frontal sulcus to the middle temporal gyrus. Findings from both DCM studies were consistent with the model proposed by Friederici (2012) and bear some similarity to the results found in the current study.

When considering the modulations induced by syntactic movement processing, the model described in Friederici (2012) explained that the backward projection from the left pars opercularis of the IFG (BA44) to the posterior superior temporal cortex is responsible for integrating semantic and syntactic information. That is, syntactic structure analysis precedes thematic role re-analysis, i.e., understanding who is doing what to whom. The left IFG, in addition, plays a role in response selection (Swick et al., 2008); in this case, comparing the semantic information expressed by the spoken sentence and the picture. It may be the case that, following integration of semantics and syntax, information is sent to the IFG in order to compare the sentential meaning to the visual information.

Within the context of the Bornkessel-Schlesewsky and Schlesewsky (2013) model, the interpretation would be that lexical-semantic and verb argument structure processing occurs in the LSTGp, followed by semantic combinatorial processes along the ventral temporofrontal pathway and syntactic combinatorial processes along the dorsal temporofrontal pathway. Pathways would converge in the LIFGop for unification of the semantic and syntactic information. However, it would then be unclear why the frontotemporal pathway is also modulated by syntactic movement.

The present results also found that connectivity between the LSFGm and LIFGop increased with both types of syntactic movement, though the role of the LSFGm is not clear. Some evidence suggests that activity in and around the LSFGm may reflect response preparation (Corbetta and Shulman, 2002; Kristensen et al., 2013) and/or cognitive control (Henry et al., 2004; Dosenbach et al., 2006). Two sentence comprehension studies previously reported activity in the LSFGm, along with the LIFG, in which the tasks involves sentence-picture matching (Kinno et al., 2008; Segaert et al., 2013). One did not provide an interpretation for LIFG activation, while the other associated it with general linguistic processing. A third sentence comprehension study found that the LSFGm was not only engaged for implausible sentences compared to plausible sentences, but also during Stroop and Flanker tasks (Ye and Zhou, 2009). Taken together, these findings suggest that the LSFGm may support domain-general cognitive processes, such as incongruence detection, as this seems to be the overarching process across the reviewed papers and present study. Within the context of the present experiment, it may be that the LSFGm is utilized for comparing the sentential meaning to the visual information, thereby allowing the participant to determine a match or mismatch response.

Finally, it should be noted that this analysis was hypothesis-driven and only included model configurations that were compatible with accounts of sentence processing supported either by Friederici (2012) or Bornkessel-Schlesewsky and Schlesewsky (2013). In other words, this study was not designed to exhaustively test all possible model configurations.

In conclusion, activation and connectivity patterns from this study were consistent with previous research supporting the model of auditory sentence comprehension posed by Friederici (2012). This model claims that processing complex sentences involves assigning grammatical relations, which is linked to the opercular part of the left IFG, followed by thematic role re-analysis, which is associated with posterior temporal cortex. Our results, however, should be taken with caution as peak fMRI activation was variable in location and strength across participants, though these factors were most consistent in the nodes of the effective connectivity models. Also, the DCM analysis did not exhaust all possible model configurations, though the model space was limited to include the most plausible model configurations according to neurolinguistics theories and to increase efficiency by optimizing computational processing time. Given these limitations, the findings from the present study suggest some greater complexity in the grammatical relationships and thematic role assignments when processing non-canonical sentences with wh-movement compared to those with NP-movement.

AUTHOR CONTRIBUTIONS

EE and CT worked together on the project from conceptualization to publication of the results. CT served as the primary scientific mentor to EE, and contributed significantly to the experimental design and methods, including development of behavioral and neuroimaging tasks. EE acquired and analyzed the MRI data and conducted the effective connectivity analyses under the direction of CT and with additional guidance from SK and DG. All authors contributed to the interpretation of results. EE took the lead in writing the manuscript and created the figures and tables. CT, DG, and SK provided critical feedback on all drafts of the manuscript.

FUNDING

This work was supported by the NIH-NIDCD, Clinical Research Center Grant, P50DC012283 (PI: CT), and the Graduate Research Grant and School of Communication Graduate Ignition Grant from Northwestern University (awarded to EE).

ACKNOWLEDGMENTS

The authors wish to thank Dr. Jennifer Mack, Dr. Elena Barbieri, Sarah Chandler, Brianne Chiappetta, and Stephanie Gutierrez for their helpful comments on the experimental design and assistance with data acquisition.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00027/full#supplementary-material

REFERENCES

Adank, P. (2012a). Design choices in imaging speech comprehension: an activation likelihood estimation (ALE) meta-analysis. *Neuroimage* 63, 1601–1613. doi: 10.1016/j.neuroimage.2012.07.027

Adank, P. (2012b). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain Lang.* 122, 42–54. doi: 10.1016/j.bandl.2012.04.014

Alario, F. X., Chainay, H., Lehericy, S., and Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Res.* 1076, 129–143. doi: 10.1016/j.brainres.2005.11.104

Alpert, K., Kogan, A., Parrish, T., Marcus, D., and Wang, L. (2016). The northwestern university neuroimaging data archive (NUNDA). *Neuroimage* 124, 1131–1136. doi: 10.1016/j.neuroimage.2015.05.060

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., and Grodzinsky, Y. (2003). The neural reality of syntactic transformations: evidence from functional magnetic resonance imaging. *Psychol. Sci.* 14, 433–440. doi: 10.1111/1467-9280.01459

Ben-Shachar, M., Palti, D., and Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage* 21, 1320–1336. doi: 10.1016/j.neuroimage.2003.11.027

Bornkessel, I., Zysset, S., Friederici, A. D., Yves von Cramon, D., and Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage* 26, 221–233. doi: 10.1016/j.neuroimage.2005.01.032

Bornkessel-Schlesewsky, I., Grewe, T., and Schlesewsky, M. (2012). Prominence vs. aboutness in sequencing: A functional distinction within the left inferior frontal gyrus. *Brain Lang.* 120, 96–107. doi: 10.1016/j.bandl.2010.06.004

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain Lang.* 125, 60–76. doi: 10.1016/j.bandl.2013.01.010

Boylan, C., Trueswell, J. C., and Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: a multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia* 78, 130–141. doi: 10.1016/j.neuropsychologia.2015.10.007

Boylan, C., Trueswell, J. C., and Thompson-Schill, S. L. (2017). Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain Lang.* 169, 8–21. doi: 10.1016/j.bandl.2017.01.008

Campbell, K. L., Samu, D., Davis, S. W., Geerligs, L., Mustafa, A., and Tyler, L. K. (2016). Robust resilience of the frontotemporal syntax system to aging. *J. Neurosci.* 36, 5214–5227. doi: 10.1523/JNEUROSCI.4561-15.2016

Campbell, K. L., and Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Curr. Opin. Behav. Sci.* 21, 132–137. doi: 10.1016/j.cobeha.2018.04.008

Caplan, D., Alpert, N., and Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *Neuroimage* 9, 343–351. doi: 10.1006/nimg.1998.0412

Caplan, D., Stanczak, L., and Waters, G. (2008). Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *J. Cogn. Neurosci.* 20, 643–656. doi: 10.1162/jocn.2008.20044

Catani, M., Mesulam, M.-M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., et al. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain* 136, 2619–2628. doi: 10.1093/brain/awt163

Chomsky, N. (1986). *Barriers, Vol. 13*. Cambridge, MA: The MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Costafreda, S. G., Fu, C. H., Lee, L., Everitt, B., Brammer, M. J., and David, A. S. (2006). A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. *Hum. Brain Mapp.* 27, 799–810. doi: 10.1002/hbm.20221

Crozier, S., Sirigu, A., Lehéricy, S., van de Moortele, P.-F., Pillon, B., Grafman, J., et al. (1999). Distinct prefrontal activations in processing sequence at the sentence and script level: an fMRI study. *Neuropsychologia* 37, 1469–1476. doi: 10.1016/S0028-3932(99)00054-8

den Ouden, D. B., Saur, D., Mader, W., Schelter, B., Lukic, S., Wali, E., et al. (2012). Network modulation during complex syntactic processing. *Neuroimage* 59, 815–823. doi: 10.1016/j.neuroimage.2011.07.057

Dick, A. S., Bernal, B., and Tremblay, P. (2014). The language connectome: new pathways, new concepts. *Neuroscientist* 20, 453–467. doi: 10.1177/1073858413513502

Dickey, M. W., Choy, J. J., and Thompson, C. K. (2007). Real-time comprehension of wh- movement in aphasia: evidence from eyetracking while listening. *Brain Lang.* 100, 1–22. doi: 10.1016/j.bandl.2006.06.004

Dickey, M. W., and Thompson, C. K. (2004). The resolution and recovery of filler-gap dependencies in aphasia: evidence from on-line anomaly detection. *Brain Lang.* 88, 108–127. doi: 10.1016/S0093-934X(03)00283-9

Dickey, M. W., and Thompson, C. K. (2009). Automatic processing of wh- and NP- movement in agrammatic aphasia: evidence from eye-tracking. *J. Neurolinguistics* 22, 563–583. doi: 10.1016/j.jneuroling.2009.06.004

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron* 50, 799–812. doi: 10.1016/j.neuron.2006.04.031

Dreher, J.-C., and Grafman, J. (2003). Dissociating the roles of the rostral anterior cingulate and the lateral prefrontal cortices in performing two tasks simultaneously or successively. *Cerebral Cortex* 13, 329–339. doi: 10.1093/cercor/13.4.329

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A.* 113:7900–5. doi: 10.1073/pnas.1602413113

Feng, S., Legault, J., Yang, L., Zhu, J., Shao, K., and Yang, Y. (2015). Differences in grammatical processing strategies for active and passive sentences: an fMRI study. *J. Neurolinguistics* 33, 104–117. doi: 10.1016/j.jneuroling.2014.09.002

Finocchiaro, C., Capasso, R., Cattaneo, L., Zuanazzi, A., and Miceli, G. (2015). Thematic role assignment in the posterior parietal cortex: a TMS study. *Neuropsychologia* 77, 223–232. doi: 10.1016/j.neuropsychologia.2015.08.025

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Grodzinsky, Y., and Santi, A. (2008). The battle for Broca's region. *Trends Cogn. Sci.* 12, 474–480. doi: 10.1016/j.tics.2008.09.001

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004

Henry, R. G., Berman, J. I., Nagarajan, S. S., Mukherjee, P., and Berger, M. S. (2004). Subcortical pathways serving cortical language sites: initial experience with diffusion tensor imaging fiber tracking combined with intraoperative language mapping. *Neuroimage* 21, 616–622. doi: 10.1016/j.neuroimage.2003.09.047

Hirotani, M., Makuuchi, M., Ruschemeyer, S., and Friederici, A. D. (2011). Who was the agent? The neural correlates of reanalysis processes during sentence comprehension. *Hum. Brain Mapp.* 32, 1775–1787. doi: 10.1002/hbm.21146

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., and Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioral and fMRI study. *Neuroimage* 44, 1152–1162. doi: 10.1016/j.neuroimage.2008.09.043

Kinno, R., Kawamura, M., Shioda, S., and Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by picture-sentence matching task. *Hum. Brain Mapp.* 29, 1015–1027. doi: 10.1002/hbm.20441

Kristensen, L. B., Wang, L., Petersson, K. M., and Hagoort, P. (2013). The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex* 23, 1836–1848. doi: 10.1093/cercor/bhs164

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *J. Psycholinguist. Res.* 33, 51–73. doi: 10.1023/B:JOPR.0000010514.50468.30

Lewis, G. A., Poeppel, D., and Murphy, G. L. (2015). The neural bases of taxonomic and thematic conceptual relations: an MEG study. *Neuropsychologia* 68, 176–189. doi: 10.1016/j.neuropsychologia.2015.01.011

Mack, J. E., Meltzer-Asscher, A., Barbieri, E., and Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain Sci.* 3, 1198–1214. doi: 10.3390/brainsci3031198

Makuuchi, M., and Friederici, A. D. (2013). Hierarchical functional connectivity between the core language system and the working memory system. *Cortex* 49, 2416–2423. doi: 10.1016/j.cortex.2013.01.007

Makuuchi, M., Grodzinsky, Y., Amunts, K., Santi, A., and Friederici, A. D. (2012). Processing noncanonical sentences in Broca's region: reflections of movement distance and type. *Cereb. Cortex* 23, 694–702. doi: 10.1093/cercor/bhs058

Marreiros, A. C., Kiebel, S. J., and Friston, K. J. (2008). Dynamic causal modelling for fMRI: a two-state model. *Neuroimage* 39, 269–278. doi: 10.1016/j.neuroimage.2007.08.019

Martino, J., and Lucas, E. M. (2014). Subcortical anatomy of the lateral association fascicles of the brain: a review. *Clin. Anat.* 27, 563–569. doi: 10.1002/ca.22321

Matchin, W., and Hickok, G. (2016). 'Syntactic Perturbation'during production activates the right IFG, but not broca's area or the ATL. *Front. Psychol.* 7:241. doi: 10.3389/fpsyg.2016.00241

Mauner, G., Fromkin, V. A., and Cornell, T. L. (1993). Comprehension and acceptability judgments in agrammatism: disruptions in the syntax of referential dependency. *Brain Lang.* 45, 340–370. doi: 10.1006/brln.1993.1050

Mizuno, K., Tanaka, M., Tanabe, H. C., Sadato, N., and Watanabe, Y. (2012). The neural substrates associated with attentional resources and difficulty of concurrent processing of the two verbal tasks. *Neuropsychologia* 50, 1998–2009. doi: 10.1016/j.neuropsychologia.2012.04.025

Nagel, H. N., Shapiro, L. P., and Nawy, R. (1994). Prosody and the processing of filler-gap sentences. *J. Psycholinguist. Res.* 23, 473–485. doi: 10.1007/BF02146686

Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090. doi: 10.1016/j.neuroimage.2007.02.041

Rogalsky, C., and Hickok, G. (2011). The role of Broca's area in sentence comprehension. *J. Cogn. Neurosci.* 23, 1664–1680. doi: 10.1162/jocn.2010.21530

Salis, C., and Edwards, S. (2005). Comprehension of wh-questions in agrammatism: a single-case study. *Read. Work. Pap. Linguist.* 8, 219–233. Available online at: http://www.reading.ac.uk/acadepts/ll/app_ling/internal/wp8/Salis&Edwards.pdf

Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2013). "Computational primitives in syntax and possible brain correlates," in *The Cambridge Handbook of Biolinguistics*, eds C. Boeckx and K. Grohmann (Cambridge: Cambridge University Press), 257–282. doi: 10.1017/CBO9780511980435.016

Segaert, K., Kempen, G., Petersson, K. M., and Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: an fMRI study. *Brain Lang.* 124, 174–183. doi: 10.1016/j.bandl.2012.12.003

Seghier, M. L., Zeidman, P., Neufeld, N. H., Leff, A. P., and Price, C. J. (2010). Identifying abnormal connectivity in patients using dynamic causal modeling of FMRI responses. *Front. Syst. Neurosci.* 4:142. doi: 10.3389/fnsys.2010.00142

Shetreet, E., and Friedmann, N. (2014). The processing of different syntactic structures: fMRI investigation of the linguistic distinction between wh-movement and verb movement. *J. Neurolinguistics* 27, 1–17. doi: 10.1016/j.jneuroling.2013.06.003

Skipper, J. I. (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Phil. Trans. R. Soc. B* 369:20130297. doi: 10.1098/rstb.2013.0297

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109. doi: 10.1016/j.neuroimage.2009.11.015

Swick, D., Ashley, V., and Turken, U. (2008). Left inferior frontal gyrus is critical for response inhibition. *BMC Neurosci.* 9:102. doi: 10.1186/1471-2202-9-102

Thompson, C. K., Bonakdarpour, B., and Fix, S. F. (2010a). Neural mechanisms of verb argument structure processing in agrammatic aphasic and healthy age-matched listeners. *J. Cogn. Neurosci.* 22, 1993–2011. doi: 10.1162/jocn.2009.21334

Thompson, C. K., den Ouden, D.-B., Bonakdarpour, B., Garibaldi, K., and Parrish, T. B. (2010b). Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. *Neuropsychologia* 48, 3211–3227. doi: 10.1016/j.neuropsychologia.2010.06.036

Thompson, C. K., and Meltzer-Asscher, A. (2014). "Neurocognitive mechanisms of verb argument structure processing," in *Structuring the Argument: Multidisciplinary Research on Verb Argument Structure*, eds A. Bachrach, I. Roy, and L. Stockall (Amsterdam: John Benjamins), 141–168. doi: 10.1075/lfab.10.07tho

Van Essen, D. C., and Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.* 6, 370–375. doi: 10.1016/0166-2236(83)90167-4

Wang, X., Song, X., Mack, J., Thompson, C. K., and Parrish, T. B. (2015). "Reliability of language network BOLD activation," in *Paper presented at the Society for the Neurobiology of Language* (Chicago, IL).

Ye, Z., and Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *Neuroimage* 48, 280–290. doi: 10.1016/j.neuroimage.2009.06.032

Yokoyama, S., Okamoto, H., Miyamoto, T., Yoshimoto, K., Kim, J., Iwata, K., et al. (2006). Cortical activation in the processing of passive sentences in L1 and L2: an fMRI study. *Neuroimage* 30, 570–579. doi: 10.1016/j.neuroimage.2005.09.066

Yokoyama, S., Watanabe, J., Iwata, K., Ikuta, N., Haji, T., Usui, N., et al. (2007). Is Broca's area involved in the processing of passive sentences? An event-related fMRI study. *Neuropsychologia* 45, 989–996. doi: 10.1016/j.neuropsychologia.2006.09.003

Conflict of Interest Statement: SK serves as a guest associate editor for Frontiers.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 18 March 2019
doi: 10.3389/fnhum.2019.00082

Predicting Known Sentences: Neural Basis of Proverb Reading Using Non-parametric Statistical Testing and Mixed-Effects Models

Bruno Bianchi[1,2]*, *Diego E. Shalom*[2] and *Juan E. Kamienkowski*[1,2]

[1] *Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias de la Computación (ICC), CONICET-Universidad de Buenos Aires, Buenos Aires, Argentina,* [2] *Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina*

Edited by:
Srikantan S. Nagarajan,
University of California, San Francisco,
United States

Reviewed by:
Mirret El-Hagrassy,
Spaulding Neuromodulation Center,
Spaulding Rehabilitation Hospital,
United States
Luis Eduardo Coutinho
Castelo-Branco,
Spaulding Neuromodulation Center,
Spaulding Rehabilitation Hospital,
United States
Jacobo Diego Sitt,
Institut National de la Santé et de la
Recherche Médicale (INSERM),
France

***Correspondence:**
Bruno Bianchi
bbianchi@dc.uba.ar

Received: *18 September 2018*
Accepted: *15 February 2019*
Published: *18 March 2019*

Citation:
Bianchi B, Shalom DE and
Kamienkowski JE (2019) Predicting
Known Sentences: Neural Basis of
Proverb Reading Using
Non-parametric Statistical Testing and
Mixed-Effects Models.
Front. Hum. Neurosci. 13:82.
doi: 10.3389/fnhum.2019.00082

Predictions of future events play an important role in daily activities, such as visual search, listening, or reading. They allow us to plan future actions and to anticipate their outcomes. Reading, a natural, commonly studied behavior, could shed light over the brain processes that underlie those prediction mechanisms. We hypothesized that different mechanisms must lead predictions along common sentences and proverbs. The former ones are more based on semantic and syntactic cues, and the last ones are almost purely based on long-term memory. Here we show that the modulation of the N400 by Cloze-Task Predictability is strongly present in common sentences, but not in proverbs. Moreover, we present a novel combination of linear mixed models to account for multiple variables, and a cluster-based permutation procedure to control for multiple comparisons. Our results suggest that different prediction mechanisms are present during reading.

Keywords: electroencephalography, reading, N400, predictability, linear mixed models, cluster-based permutation test

1. INTRODUCTION

When performing any task, such as visual searches, listening, or reading, the brain is not passively waiting to be activated by external stimuli. Instead, it is actively trying to predict those upcoming events, planning future actions and anticipating their outcomes (Kveraga et al., 2007). Reading, a natural, commonly studied behavior, could shed light over the brain processes that underlie those prediction mechanisms. In the early 80s, Kutas and Hillyard (1980) observed for the first time a late negative deflection that appeared 400 ms (N400) after the onset of semantically incongruent words compared with congruent words (e.g., "I take coffee with cream and **car**" compared with "I take coffee with cream and **sugar**"). They then described that the amplitude of this N400 was correlated with the *Cloze-Probability* or *Predictability* (i.e., the proportion of subjects that fill in a particular word as the most probable next word in a sentence). They concluded that not only semantic incongruities elicited this negative deflection, but so also did words that had low Predictability (e.g., "I take coffee with cream and **cinnamon**" versus "I take coffee with cream and **sugar**") (Kutas and Hillyard, 1984).

Two classical views of the N400, the integration interpretation and the lexical-access interpretation, are found in the literature. The integration view proposed that *"it reflects the process of semantic integration of the critical word with the working context"* (Lau et al., 2008). The lexical view suggested that *"it reflects facilitated activation of features of the long-term memory (LTM) representation that is associated with a lexical item"* (Lau et al., 2008). Both views easily explained the gradual nature of the N400: in the former view, because they were more difficult to combine with the previous context and, in the latter, because higher associated word-contexts implied lower difficulty, elicited smaller responses, and was independent of the combination/integration process that updated the context. Some years later, Kutas and Federmeier (2011), proposed a middle-term interpretation, in which the N400 represented the process of binding the current long-term memory (LTM) landscape with the incoming new stimulus. Thus, a large activity was produced when the incoming stimulus mismatched the current landscape. Moreover, the amplitude of this activity was modulated by the degree of mismatch or, inversely, the Predictability of the incoming word (Lau et al., 2008).

Nowadays, most neurolinguistic experiments on predictions use sentences with simple contexts [e.g., "I take coffee (...)"], as in Kutas' first studies. When these statements were presented, it was hypothesized that a subset of words within the semantic field was pre-activated (Lewis et al., 2006) (e.g., "cup", "sugar", "toast", "cream"). But, these "semantic predictions" were not enough for engaging accurate predictions. For instance, in the previous example, the activated words were nouns, but following the rules of English, the statement "I take coffee" cannot continue with another noun. Thus, to generate well-formed sentences, it is necessary to also make "syntactic predictions" (e.g., a preposition like "with") (Boston et al., 2008). In addition, there are scenarios in natural reading where we find previously known sentences, like in the so called multi-word strings (e.g., idioms, proverbs, song lyrics) (Vespignani et al., 2010; Molinaro et al., 2013). The predictions performed on these sentences are "mnemonic predictions" and, despite the fact that these are found commonly in everyday language, they are largely unexplored in the literature.

The main difference in processing these memory-encoded sentences compared with common sentences is that, in the former, there is a moment where the linguistic context (i.e., the sum of previous words) triggers the recall of the rest of the sentence. Therefore, the upcoming words become highly predictable regardless of whether they are syntactically incorrect or semantically unrelated. That point was called "Recognition Point" (RP: a word that enables the reader to recognize the read sentences) by Vespignani et al. (2010) and "MaxJump" (MJ: a word with the maximal difference in Predictability with the previous word) by Fernández et al. (2014).

To our knowledge, there are very few studies on these memory-encoded sentences, and they are mainly focused in memory-encoded structures within sentences. In several studies, Molinaro et al. explored the first and last words of idioms (e.g., *"break the ice"*) (Molinaro and Carreiras, 2010; Vespignani et al., 2010; Molinaro et al., 2013) and the last word of complex prepositions (i.e., *"in relation to"*) (Molinaro et al., 2008). For instance, in their experiment, they found a larger N400-like component in expected final words compared with unexpected final words (Molinaro et al., 2008), which suggested that Cloze-Probability did not capture all the variables involved in prediction processes. Moreover, two separate late responses were present after the last word of the idiom: a P300 that resembled the expectancy of that word and a N400 sensitive to the semantic properties of that word (Molinaro and Carreiras, 2010). This was further supported by analyzing the first word of the idiom, which generally matched the RP. At that word, semantic violations but not substitutions elicited a N400. In contrast, in the following word, where the context was already known, both elicited a N400 (Vespignani et al., 2010). Moreover, changes in theta and gamma bands and an early increase in fronto-occipital interactions in both frequency bands were observed after the RP and before the final word. That suggested that internal knowledge supported low-level, perceptual processing during reading (Molinaro et al., 2013; Monsalve et al., 2014).

Recently, Fernández et al. (2014) extended the study of multi-word strings to fully memorized sentences, such as proverbs, where they focused on the Predictability effects using eye tracking measures. They found differences in the pattern of fixations between proverbs and common sentences after the MJ (or RP). In accordance with EEG results, these differences were interpreted as a change in the prediction pathways after the recognition of the proverb. This was, to our knowledge, the only study that has explored those sentence types.

In the present work we aimed to find and distinguish brain sources of prediction mechanisms (i.e., semantic, syntactic, mnemonic) in dense EEG signals, when reading different sentence types. With that objective in mind, we focused on analyzing proverb reading and how Cloze-Task Predictability affected word processing, taking into account multiple variables. Since proverbs comes from everyday language, and manipulations would break the memory recall, our corpus consisted in stimuli that were unbalanced in several variables, such as position of the Recognition Point, Predictability, word frequency, and sentence length. Classical hypothesis testing could not cope with these multiple unbalanced co-variables, missing values, or they would require a larger amount of data. To solve these issues, we implemented a Linear Mixed Model (LMM) for each sample (time-point and electrode), which allowed us to test several categorical and continuous co-variables at once. A downside is that this resulted in too many comparisons (i.e., as much as the number of electrodes by time-points). The Cluster-based permutation (CBP) procedure is a non-parametric statistical method that corrects for these multiple comparisons from the sample-by-sample testing in M/EEG data (Oostenveld et al., 2011). This procedure is widely used in the field nowadays, mainly in combination with *t*-test. In this work we replaced the sample-by-sample *t*-test by multivariate LMMs. The combination of these techniques (LMM-CBP) offers a powerful statistical test for M/EEG analysis.

2. MATERIALS AND METHODS

2.1. Subjects

Twenty-eight healthy participants took part in the experiment [(24.3 ± 4.2) years old; 12 females], receiving monetary reward for their participation. Three subjects were excluded from analysis due to noisy signal acquisition. Every session took 1.5–2.0 h, which included preparation. All participants provided written informed consent in agreement with the Helsinki declaration, and they were reimbursed monetarily for their participation after the study. All the experiments described in this paper were reviewed and approved by the ethics comittee: "Comité de Ética del Centro de Educación Médica e Investigaciones Clínicas "Norberto Quirno" "(CEMIC)" and qualified by the Department of Health and Human Services (HHS, USA): IRb00001745 - IORG 0001315 (Protocol 435).

2.2. Task

Each trial consisted of an entire sentence (example in **Figure 1A**) presented word by word in the center of the screen. Every word was presented for 300 ms, with an Inter-Stimulus Interval (ISI) of 400 ms (**Figure 1B**; SOA = 700ms). The entire trial duration depended on the sentence length (min = 5 words, max = 12 words). Before starting, participants performed 10 trials of training that were not analyzed. Participants were instructed to concentrate on the sentences and to avoid eye movements during trials.

After practice, participants performed 120 trials that were divided into four blocks with unlimited time for the participant to rest between them. A fixation cross lasting 1 s in the center of the screen indicated the start of the trial. Within blocks, the inter-trial interval was 2 s. To ensure that subjects were reading consciously, they answered a simple multiple choice question every six trials (randomized), on average. There was no timeout for answering, and participants were instructed to use this lapse for eye resting. Almost all the responses were correct in all participants (accuracy = 98%).

Visual stimuli were prepared and presented using Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). They were shown on a 19-inch CRT monitor at 60 cm from the participant's eyes at a refresh rate of 60 Hz.

2.3. Sentence Corpus

For the present work we selected a subset of 130 sentences from the corpus used by Fernández et al. in a study of eye movements (Fernández et al., 2014). They estimated word-Predictability for every word, using a Cloze-Task (Taylor, 1953), with 18 graduate and undergraduate students. In this task, subjects had to fill the most probable word following an incomplete sentence. Word-Predictability was then estimated for each word as the proportion of correct answers. Our subset included 50 memory-encoded and 80 common sentences, and 5 sentences of each category were used in the training stage. For the memory-encoded sentences, it was possible to find a Recognition Point or Max Jump (Fernández et al., 2014) for each sentence (**Figure 1C**).

The corpus consisted of 897 words (470 content words) with 451 unique words (368 unique content words). The median logit Predictability for memory-encoded sentences was 0.186 ± 0.948, and for common sentences was -0.430 ± 0.897 (**Figure 1D**).

2.4. EEG Recording and Preprocessing

Electroencephalography (EEG) signals were recorded using a Biosemi Active-Two (Amsterdam, The Netherlands) 128-channel system at 1024 Hz. All the analyses were performed using EEGLAB (Delorme and Makeig, 2004; Makeig et al., 2004), FieldTrip (Oostenveld et al., 2011), and in-house MATLAB and R scripts. Data were re-referenced to linked mastoid. A Hamming-windowed FIR band-pass filter of 0.1–40 Hz was applied, using "eegfiltnew" in EEGLAB v14.1 (Widmann and Schröger, 2012), and data were downsampled to 256 Hz. For the Event-Related Potential (ERP) analysis, data were epoched from 100 ms before to 700 ms after the onset of the stimulus, with the amplitudes from -100 ms to the onset as the epoch baseline. Epochs with > 5 electrodes with at least one sample of 80 μV were rejected. Ocular artifacts were detected using Independent Component Analysis (ICA) and removed after manual inspection of the components for typical ocular topography.

2.5. Analysis and Statistics

We implemented four types of analyses (**Table 1**), which are described in detail below. Briefly, some approaches use *a priori* defined ROIs and time-window to extract a single value for each trial. Then, it is possible to discretize variables, average across categories, and apply different hypothesis tests –such as Kruskal-Wallis or Wilcoxon's tests–, or to preserve the continuous variables and apply a regression analysis. Nevertheless, these approaches imply a huge loss of information in the averaging procedure. In order to avoid that, it is possible to run a test in each sample (electrode and time-point) and deal with the multiple comparisons problem using for instance a cluster-based permutation (CBP) procedure. This procedure is widely used in multichannel recordings, since it takes into account the high correlations between channels with little loss of power, as opposed to Bonferroni or false-discovery rate approaches (Maris and Oostenveld, 2007).

2.5.1. Categorical Predictability in the N400-window

Only content words (adjectives, verbs, and nouns) were kept for all the analyses. Words shorter than three characters, and the first word of each sentence were also rejected. All the words were classified independently in terciles according to the values of both Frequency and Predictability of the corpus. Evoked potentials were averaged in the chosen ROI (**Figure 2**, central inset), time-window (between 300 and 450 ms), condition (tercile of Frequency or Predictability, and Sentence Type) and participant.

The average for each participant in the N400-window were submitted to a Paired Wilcoxon Rank-Sum Test with Sentence Type (two levels) as main factor, and to two independent Kruskal-Wallis Tests with Predictability (three levels) and Frequency (three levels).

2.5.2. Continuous Predictability in the N400-window

A Linear Regression was performed to address for the continuous Predictability effect (using its raw numerical value) on the

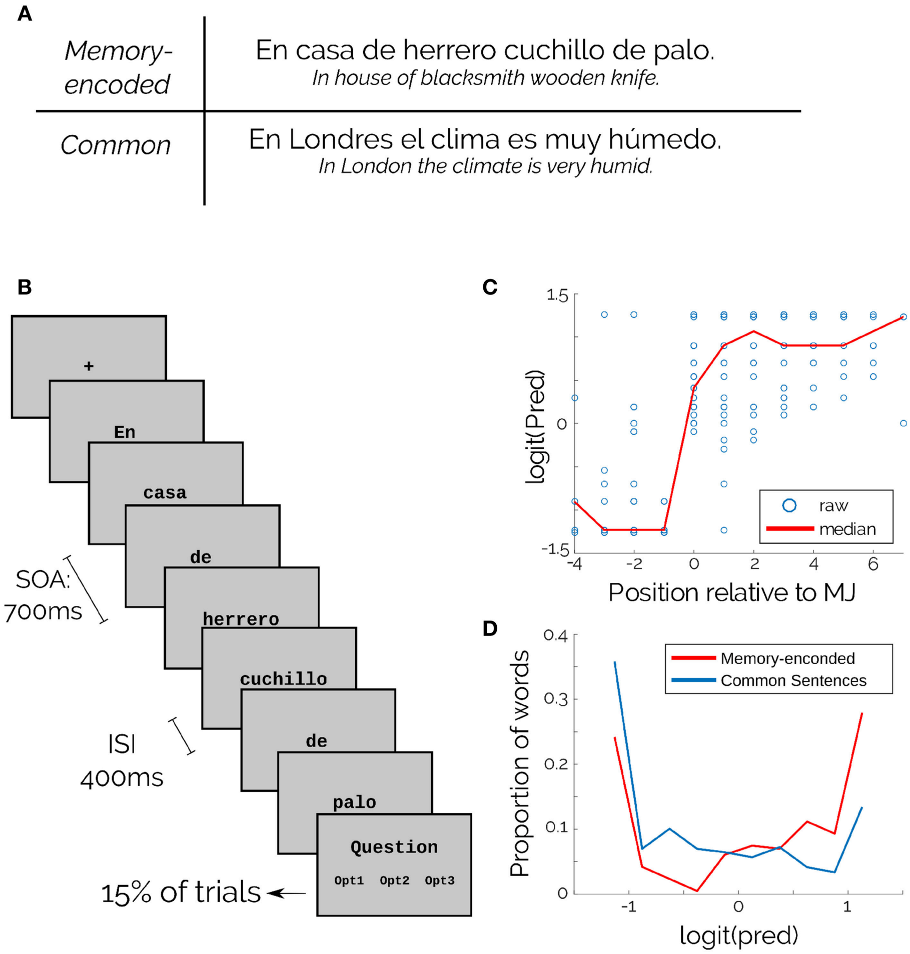


FIGURE 1 | Experimental design and materials. **(A)** Examples of both types of sentences. **(B)** Schematic of the task. **(C)** Word Predictability of memory-encoded sentences. Position is relative to the RP. Blue dots are individual words and red line is the median. **(D)** Predictability of memory-encoded (Red) and common sentences (Blue).

N400-window. The Linear Regression was fitted using single-trial values from all the participants together, with participants included as dummy variables.

2.5.3. Categorical Predictability of the Whole Epoch

In order to detect predictability effects occurring on the whole epoch –i.e., on any time point or electrode–, a common solution is to apply a CBP test, in which averaging across electrodes and time points is not required. The current Fieldtrip's implementation supports mainly categorical factors (Maris and Oostenveld, 2007; Oostenveld et al., 2011). This toolbox was used to run two separate tests for Predictability (two levels: High vs Low) and Sentences Type (two levels).

2.5.4. Continuous Predictability of the Whole Epoch

The use of any type of natural sentence that appears in everyday vocabulary implies using a non-uniform corpus of sentences; frequency, length, Predictability, and other word properties cannot be controlled and balanced across trials. This is an important issue when using classical ERP analysis, because when averaging across one condition other conditions may become unbalanced. In recent years, computational advances have allowed researchers in the neurolinguistics field to handle this issue using linear regressions, both in the study of eye movements and in ERP analysis. In the former case, Linear Mixed Models (LMM) have become the most common methodological technique, because they allow for testing multiple co-variables at once, and they also account for random effects (e.g., of subjects and items) that are necessary to avoid the language-as-fixed-effect fallacy in studies that involve natural language (Clark, 1973). LMMs are not widespread in the ERP field. To our knowledge, the only implementation of LMM applied to ERPs is in the LIMO toolbox for MATLAB (Pernet et al., 2011), but it is focused on "assessing the inter-subject variability" and not for testing for effects of co-variates. Finally, other classical regression analysis have just been introduced in the last few years (Hauk et al., 2006; Smith and Kutas, 2015a,b), but without the benefits of LMMs and CBP pocedure stated above.

TABLE 1 | Statistical approaches on analyzing EEG data for the Predictability effect (or any continuous variable).

	N400 window	Complete epoch
Discrete Predictability	Kruskal-Wallis and Wilcoxon	CBPT (*t*-test)
Continuous Predictability	Regression	LMM-CBPT (NEW)

After estimating the effects for each co-variable, another important benefit of fitting LMMs is that it is possible to extract predictions (partial effects) from the original data. The *remef()* package for R (Hohenstein and Kliegl, 2013) takes the results of a LMM as input and uses them to remove the partial effects for those co-variables from the original data. Analyzing this newly generated data allows us to understand further the remaining cleaner effects.

In the present work, LMMs were fitted using lme4 package V1.1-12 (Bates et al., 2015b) for R V3.3.2 (R Development Core Team, 2008) as follows: each time (t) and electrode (e) sample of all the epochs were used as dependent variables. A combination of co-variables (and some interactions) were used as independent variables in the following model:

$$Amp_{(t,e)} \sim Freq + Pred : Type + Pos : Type + (1|Subj) + (1|Word) \tag{1}$$

where $Freq$ is the Frequency on the lexicon, $Pred$ is the Predictability, $Type$ refers to the Sentences Type, and Pos is the Ordinal Position in the sentences. Subject ID ($Subj$) and the string of each word ($Word$) were used as random factors. The colon between two co-variables indicates that we are testing the interaction between them. The relevant output of these models is an estimate of the slope and its error (SD) for each of the fixed factors in Equation 1. With this information it is possible to calculate a t-value, which represents how far away from zero the estimate is. Commonly, if the estimate is more than 2.0 SD away from zero, the slope is considered significant with $\alpha < 0.05$ (Fernández et al., 2014).

Because this model was fitted for each electrode-time sample, the final results of the analysis were n-by-m t-values matrices (i.e., one matrix with the t-values of each model run for each term of the model), with n electrodes (128) and m time samples (103). This means that more than 10,000 t-values were in consideration at once, which is a huge number of comparisons that needs to be corrected for multiple comparisons to control for Type I error. To solve this multiple comparison issue without losing statistical power, we implemented a CBP protocol proposed by Maris and Oostenveld (Maris and Oostenveld, 2007), and we adapted it to the LMMs.

2.5.4.1. Permutation procedure

Maris & Oostenveld (Maris and Oostenveld, 2007) introduced the CBP procedure that proposed a novel way to analyze EEG data in a non-parametric framework, which avoided *a priori* hypotheses of time and scalp distribution. This procedure consists of running a statistical test for each electrode-time sample.

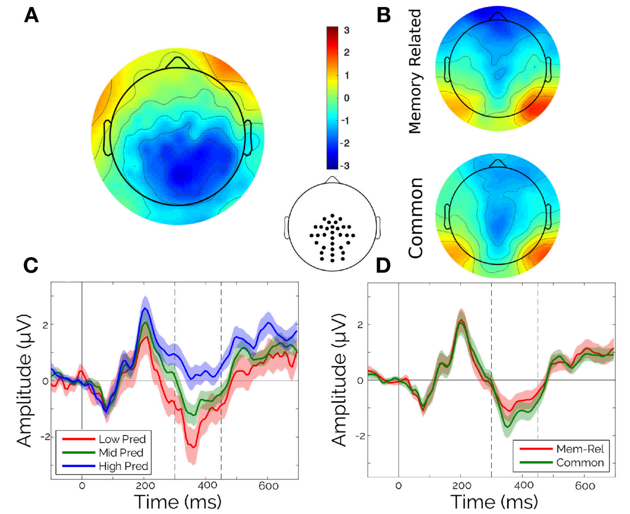


FIGURE 2 | Classical ERP analysis. **(A)** Scalp topography of the difference between High and Low terciles of Predictability in the [300 450]ṁs interval. **(B)** Scalp topography for memory-encoded (top) and common (bottom) sentences, in the [300 450]ṁs interval. **(C)** Average waveforms for a centro-parietal cluster (see inset) separated by terciles of Predictability. **(D)** Average waveforms for a centro-parietal cluster (see inset) separated by sentence type. Vertical lines indicate the [300 450]ṁs interval.

The CBP procedure includes the following steps:

1. Perform a statistical test for every electrode-time sample amplitude.
2. Select all samples that have a t-value larger than some t_{th} threshold.
3. Cluster the selected samples in connected sets on the basis of temporal and scalp distribution adjacency (at least two neighbor samples, electrodes and/or timepoints).
4. Calculate a cluster-level statistics (e.g., by taking the sum of the t-values within a cluster).

To define a significant cluster, a permutation procedure was used. The labels of categories of the trials were shuffled randomly and the previous procedure was repeated (steps 1–4) for each permutation p. Then, the largest cluster of each permutation was selected, and all the t-values within this cluster were summed ($MaxSum_p$). The $MaxSum_p$ values of many permutations (in our case $N_P = 500$) were collected to build a distribution. The sizes of the original clusters ($t_{cluster}$) were compared to this distribution of $MaxSum_p$. Then the p-values for each original cluster were estimated as the proportion of $MaxSum_p$ that exceeds $t_{cluster}$, over the whole set (P) (Equation 2). In the case that none of the permuted datasets exceeds the original data, the p-value is defined as less than 1 over N_P (Equation 2).

$$\begin{cases} \text{if} \quad \sum_{p \in P} (MaxSum_p > t_{cluster}) > 0, & p = \sum_{p \in P} (MaxSum_p > t_{cluster}) / N_P \\ \text{if} \quad \sum_{p \in P} (MaxSum_p > t_{cluster}) = 0, & p < 1/N_P \end{cases} \tag{2}$$

Finally, an alpha level is defined to determine the significance level of these clusters as in parametric testing.

In the case of a t-test, a single t-value is obtained from each test. Hence, a single distribution of cluster sizes (sum of t-values) was built from the permutations, and the size of the original clusters wwere compared to this distribution. In the case of LMMs statistics, one t-value was obtained for each fixed effect included in the model. Each of these values was treated separately. A single distribution of cluster sizes was obtained for each fixed effect, and the sizes of the clusters of the original models were compared with these distributions.

The main problem faced in the adaptation of a CBP protocol to LMM statistics was the multiple co-variables that need to be considered at the same time. In a multivariate analysis, where there are correlations between the co-variables, shuffling only one label would break the correlations between all the covariates. To avoid this, the trial label was shuffled, and the entire structure of correlations was keep intact. That is, each EEG matrix (i.e., the matrix with the EEG amplitudes) was assigned to a new co-variable vector (i.e., the vector with all the co-variables from a trial).

Another important problem to solve in the permutation of trials under a LMM is that permuting across the random variables breaks the random factor structure, wich generates anti-conservative results. To address this issue, we mimicked the F1/F2 approach used in psycholinguistics when fitting ANOVA models for more than one random variable (i.e., words and subjects). Here, we permuted within each of the random factors (**Figure S1**). That is, for our model, which had two random factors (i.e., subject and word), we first ran a complete CBP procedure that kept the structure for Subject and, second, we ran it again keeping the structure for Word. In the results we present both of these analyses.

2.5.4.2. Implementation

Because the major literature in LMM is based on the lme4 library for R, to implement this CBP procedure it was necessary to export data from MATLAB structures to CSV files. To facilitate the parallelization, data were exported in many CSV files, one per time sample, with all the information on amplitudes, co-variables, and random effects for each electrode. The in-lab parallelization was made using 26 4-core, *Core i7* Desktop computers (104 independent cores) at the same time to fit all the models. It took 6 min to fit all the permutations for one electrode and one time sample ($N = 500$), which made it possible to run all the models (i.e., 128 electrodes in 103 time-samples) in 12 h. For model fitting, RAM memory usage was negligible in relation to the processing cost. Additionally, it is important to remark that for each permutation, all the electrode-time samples received the same shuffling, which was pre-calculated and stored in each core of the 26 computers. The code for this analysis is available at http://reading.liaa.dc.uba.ar. It includes scripts written in MATLAB, R, and Bash, example data, and a tutorial.

3. RESULTS AND DISCUSSION

3.1. Predictability and Sentence Type Effects: Classical Approaches

As a first step, we aimed to assess the main effects of Word Predictability, Sentence Type, and Frequency in the evoked responses. Particularly, guided by the literature on Predictability effects, we looked for Predictability effects in the N400 window ([300 ms, 450 ms]; and selected electrodes, see **Figure 2**, central inset). Within this N400 window, only Predictability showed a significant effect (**Figures 2A,C**, Kruskal-Wallis: $\chi^2 = 9.02$, $p = 0.011$). Frequency (Figure not shown, Kruskal-Wallis: $\chi^2 = 1.38$, $p = 0.50$) and Sentence Type (Wilcoxon: $z = 1.69$, $p = 0.091$) showed no significant effects (**Figures 2B,D**).

However, it was possible to identify two main limitations of this type of categorical ERP analysis: on one side, for the Predictability Effect it is necessary to rely on a categorization of this continuous variable. On the other side, averaging across electrodes and time samples implies introducing *a priori* hypotheses about the effect distribution. This is critical when expecting potential effects from other co-variables. For instance, the N400 window was mainly motivated by the Predictability effect, but Frequency and Sentence Type, as well as Predictability itself, could, in principle, have effects outside this window. To overcome the first limitation, a linear regression was fitted for the numerical (logit) Predictability against the mean amplitude on the N400 window, with Subjects as a dummy variable. Extending the previous analysis, we observed significant effect of Predictability on the N400, but the variance explained by this factor was very low ($\beta = 1.1567$, $R^2 = 0.0144$, $p < 0.01$; **Figure S2A**).

In order to avoid averaging across electrodes and time points, and to overcome the second limitation, a non-parametrical cluster-based analysis was performed for both the categorical Predictability (two levels, High vs Low) and the Sentences Type. Predictability effect (**Figure S2B**) showed one significant cluster ($p = 0.002$), while the Sentences Type did not show any significant effect. The significant cluster of Predictability appeared approximately at the expected latency and location, i.e., between 300 and 450 ms over the centro-parietal electrodes. However, the present analysis enabled us to explore the dynamics of significant electrodes in more detail, in particular it was a little bit earlier (210 to 420 ms) than the a priori defined window, covering the maximum number of electrodes at 335 ms (**Figure S2C**).

3.2. Predictability, Sentence Type and Position Effects: LMM-CBP Approach

In the previous subsection we introduced two alternative analyses that independently solved the limitations found when analyzing the Predictability Effect. On one side, regressions allowed to use continuous variables. On the other side, CBP procedure with t-test allowed to avoid making *a priori* hypothesis about latency and scalp distribution, which was particularly relevant when analyzing unexplored effects, like Sentences Type. In the following, we combined Linear Mixed Models (LMMs) with a CBP procedure (LMM-CBP), in order to solve all these limitations in a single procedure.

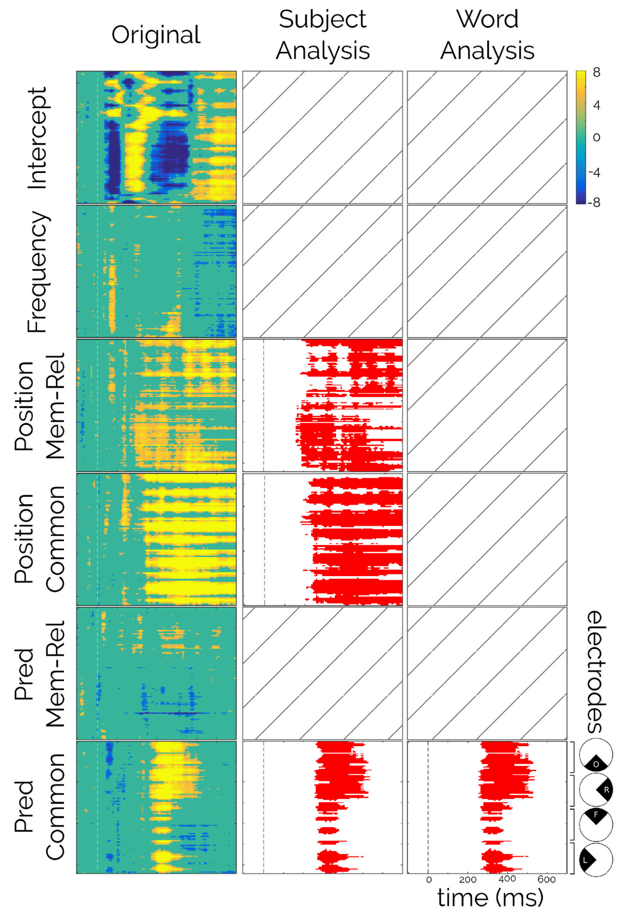


FIGURE 3 | Linear Mixed Models results. *t*-values for each fixed factor (left column), significant positive *t*-values (> 2) are presented in orange-yellow colors, significant negative *t*-values (< -2) are presented in blue colors, and non-significant *t*-values are presented in green. Significant clusters after CBP procedure for subject analysis (center column) and for word analysis (right column). Crossed squares represent no significant clusters for that effect after the correction. Rows correspond to the fixed effects: Intercept, Frequency, Position × Mem-Rel (Memory-related), Position × Common (Common words), Pred (Predictability) × Mem-Rel (Memory-related), and Pred (Predictability) × Common (Common words). Columns correspond to *t*-values from the LMM (original), and the results from Cluster-Based Permutation procedure using Subjects or Words.

The first step in the LMM-CBP procedure was to run the statistical model (Equation 1) for each time-electrode sample. Each model was fitted with 9,459 epochs of 25 participants. The results were summarized in one matrix for each fixed effect of the model (**Figure 3**, first column). These matrices showed only the significant *t*-values before the multiple comparison correction (i.e., in blue, $t < -2$; in yellow $t > 2$).

As explained in the Methods Section, the permutation of the labels was performed in two parts. First, we kept the subject structure intact (**Figure S1B** and **Figure 3**, second column) and second, we kept the word structure intact (**Figure S1C** and **Figure 3**, third column).

First, the intercept term, that is, the ERP amplitude when all the co-variables equals their own mean, resembled the usual evoked responses to visual stimuli (**Figure S3**, an early negativity N1, followed by a positivity P2, etc.) irrespective of their frequency, Predictability, position, and context. This result was expected since the Intercept should equal the mean accross all conditions (Smith and Kutas, 2015a), and it allowed us to highlight the power of this novel method to capture relevant effects in ERP signals (**Figure 3**, first row).

Second, a significant effect of the word position was observed for both types of sentences that was only significant for the subject permutation scheme (**Figure 3**, third and fourth rows, $p < 0.002$). This spatially widespread effect showed a latency of 170 ms (memory-encoded: 165 ms and common: 180 ms) and lasted the entire epoch (**Figure 4A**). Moreover, this effect seemed to be present regardless of the Predictability effect (**Figure 4B**). We attribute the absence of this effect in the word permutation scheme to the composition of the sentence corpus, where each word was presented only a few times across sentences (mean [range] = 2 [1 12]), and most of them appeared only once in the corpus (345 out of 448). Thus, when permuting within words, the ERP space-time matrix of a given word and a given subject was assigned to the same word of another subject (only changing the subject ID). This led to a very conservative criterion, that only significant effects, like the Predictability on common sentences (see below), were able to attain.

Finally, the Predictability analysis presented a significant cluster, with both permutation schemes (**Figure 3**, bottom row, $p < 0.002$ and $p = 0.005$ respectively) between 258 and 540 ms. Although widely spread across the scalp, this effect was stronger in the centro-parietal region and resembled the distribution of the N400 (**Figure 4B**). Importantly, this effect was only seen for common sentences and not for memory-encoded ones (**Figure 3**, fifth row).

The LMM analysis allowed us to make a distinction between the contributions of different co-variables. Furthermore, it makes possible to remove some of those partial effects that acted as confounding factors for the effect of interest. Initially, the huge Position effect overlapped the Predictability effect (**Figure 5A**). But, after estimating the partial effects, we were able to isolate and remove the effects of Position for both sentence types and study clean waveforms for the Predictability and Sentence Type interactions. As expected from the results from **Figure 3**, we observed a clear difference between High and Low predictable words for the common but not for the memory-encoded sentences in the N400 time-window (**Figures 3**, **5B**). Interestingly, the N400 amplitude for memory-encoded sentences was closer to the High predictable than to the Low predictable words (**Figure 5B**), suggesting the absence of N400 for either high and low predictable words in the memory-encoded sentences.

4. CONCLUSIONS

The aim of the present work was to investigate the electrophysiological basis of different prediction sources in

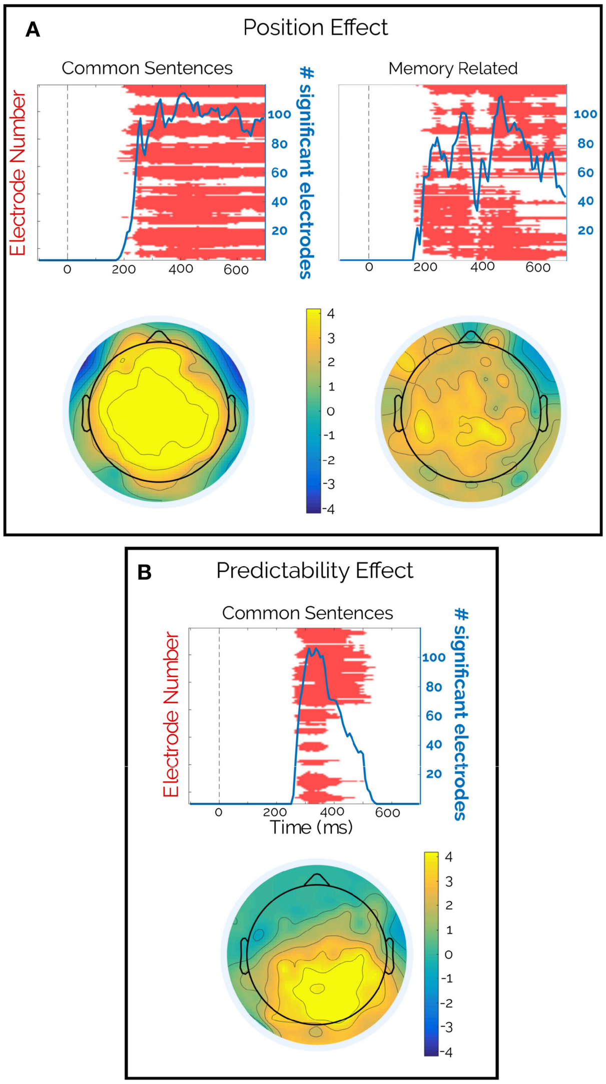


FIGURE 4 | Scalp and time distributions of significant effects. **(A)** Significant clusters (red) and cluster size (blue) as function of time (top), and topographical distribution of the *t*-values in the [300 500] ms interval (bottom) for the Position effect for both sentence types. **(B)** Significant clusters (red) and cluster size (blue) as function of time (top), and topographical distribution of the *t*-values in the same [300 500] ms interval (bottom) for the Predictability effect in common sentences. Only the subject permutation scheme is presented, because the Word permutation did not have significant clusters for the Position effect, and it had a very similar cluster in the Predictability effect.

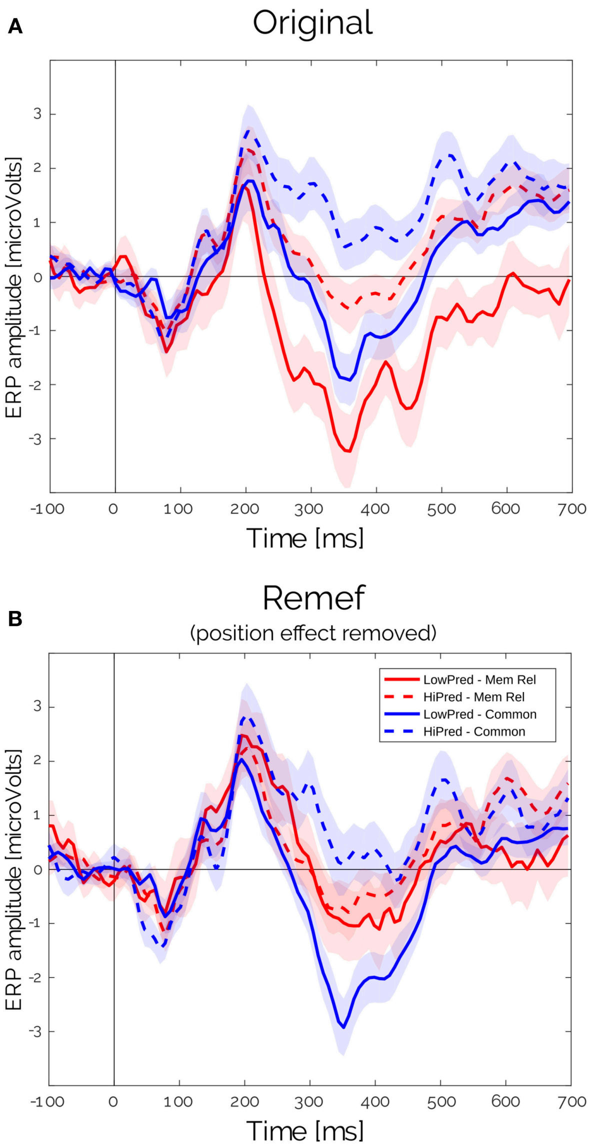


FIGURE 5 | ERP for Predictability × Sentences type for **(A)** original EEG data and **(B)** after removing the predicted position effect, as obtained from CBP procedure.

reading. To achieve this, we studied mnemonic (i.e., recalling a sequence of words from long term memory), semantic (i.e., based on the sentence topic), and syntactic predictions (i.e., predictions on syntactical rules). We used a variety of memory-encoded sentences from popular culture, like proverbs, song titles, or parts of song lyrics. All these sentences have a recognition point (RP). By using this type of natural stimuli we faced an important imbalance in the variables, such as the position of the RP, the Predictability and sentence length. This imbalance generates several difficulties for some ERP analyses commonly used, in which variables are studied one at a time, averaging across all the residual variables. In addition, these analyses require variables to be categorical –loosing information in the case of continuous variables–, and to collapse the EEG time samples and electrodes in a single value. To overcome these limitations, two partial solutions are usually applied: a linear regression, that allows modeling with continuous variables; and the Cluster-Based Permutation (CBP) procedure with *t*-test as the sample-by-sample test (Maris and Oostenveld, 2007), that allows studying brain potentials without any *a priori* hypothesis on the latency and the spatial location. Nevertheless, despite the results we obtained from these analyses were in line with the N400 bibliography, they were not a conclusive solution for those limitations, as they couldn't tackle all the limitations at once.

Taking into account those considerations, we developed a novel analytical approach. In the rERP framework proposed by Smith and Kutas (2015a), they used multivariate regressions for each sample of the epoch matrix to separate the spatial and temporal dynamics of each effect. They ended up with one time series of beta values for each effect, which was the core of their proposal. But, after that, they analyzed betas as a typical –but cleaner– ERPs (Smith and Kutas, 2015b). Here, we showed that is possible to use Linear Mixed Models (LMM) rather than regressions to generate better modeling of the data (Baayen, 2008; Bates et al., 2015a,b). And, instead of going back to the classical statistical approaches, we take advantage of the powerful statistics generated by the LMMs for each sample, and combined them with the CBP procedure introduced by Maris and Oostenveld (Maris and Oostenveld, 2007). This procedure enabled us to analyze the significance of the model slopes with a non-parametrical test for solving the multiple comparison issue generated when comparing many statistical results.

Two additional challenges were solved for our implementation: (1) the processing cost; and (2) the decision on how to perform the flag permutation of trials, as the trials could be grouped by subjects or words. The former was worked out by parallelizing the analysis. We run this in a cluster but it also could be done in a single multi-core computer. And the last one was solved by permuting in two stages –i.e., performing "subject" and "word" analysis separately–, based on the F1/F2 analysis that was traditionally used in linguistics before the popularization of LMM.

Interestingly, the proposed LMM-CBP procedure was able to model the raw ERP in the intercept term, tearing apart all the modeled effects. This is expected from the mathematical derivation of the rERP framework done by Smith and Kutas and, although it doesn't add new information, it serves as clear validation of the method (Smith and Kutas, 2015a,b). Furthermore, based on similar ideas, we were able to separate the effects of Predictability and Word Position in Sentences, which overlap on latency and scalp distribution.

The main result of this non-parametric analysis was a clear and significant effect of the word Predictability on the N400 time window, but only for the common sentences in both permutation scenarios of subjects and words. The N400 effect strongly arose without using *a priori* hypothesis of latency or localization, and it was not present in memory-encoded sentences. Interestingly, this suggests that there was neither facilitation nor a combinatorial process that relied on the previous context. This could be the case if proverbs were actually loaded from memory as a whole construct that is recalled after the recognition point was read. Moreover, the activity during the N400 period at the same centro-parietal cluster of electrodes was smaller for memory-encoded than for common sentences, as it was shown when removing the Position effect. This is in line with the results of Molinaro et al. who observed a larger negative activity in the N400 time window for different substitutions compared with original idioms or collocations (Molinaro and Carreiras, 2010). Importantly, instead of comparing between sentence types, in our case, we evaluated the effect of Predictability within each type, and no substitutions were used. Our effects, although smaller, allowed us to use the gradual nature of the N400 as a hallmark. These results suggested that the N400 was not only smaller in proverbs, but it was also insensitive to Predictability.

In addition to this N400-like effect, we observed a clear position effect that was spread widely in the scalp, which started after 200 ms of word onset. This positive drift in the EEG signal as the reader moved through the sentence was present for both sentence types. This effect could be related to a cumulative integration process. Future approaches need to parametrize information along the sentence to be included in LMMs and to relate these effects to a specific cognitive model (de Lange et al., 2010; Brouwer and Hoeks, 2013; Kamienkowski et al., 2018).

It is important to note that the position effect was only significant in the subject analysis. This could be due to the fact that when permutations were generated for the subject analysis, all the epochs of each subject were permuted within the same subject. Because we had 354 ± 76 valid epochs (after filters, see methods) per subject, we linked each ERP matrix to a different word, but, always from the same subject. Thus, the new "random" analysis broke the word structure, which generated random results for this item. Conversely, in the word analysis the permutation was done across the words (using the unique strings as the "word" levels). This means that if a word appeared only once in the entire corpus (which most of the words did), permutations were assigned an ERP matrix from one word of one subject to the linguistic information of that same word, but from another subject with a probability of $\sim 96\%$. This resulted in similar results in the permutation analysis as in the original analysis and to a high MaxSum statistic for the cluster selection, rejecting all the clusters in the original data. Thus, it was not possible to draw reliable conclusions from the Word Analysis.

Nevertheless, beyond the methodological discussion, the position effect in the subject analysis suggests a potential cumulative process during sentence reading. This effect could be separated from the cloze-task predictability effect with the present approach. In order to further analyze the slight differences in scalp distribution and strength of the position effect a follow up study could be designed using longer sentences, aligned by their Recognition Point. This would decrease the number of confounding effects.

In the present work, we present a novel analysis by combining Linear-Mixed Models and a cluster-based permutation procedure. The former are becoming very popular in eye movement and reading studies to cope with multiple, continuous, independent variables. The latter is very popular in EEG analysis and it is used to deal with the usual multiple-comparisons problem in the high density EEG signal. Using the novel LMM-CBP technique, we showed that different mechanisms are involved in the prediction of forthcoming words. Future experiments should investigate these mechanisms further to describe the precise brain areas involved and the contributions of timing and frequency, to then integrate them with cognitive models of the role of prediction in processing natural language.

DATA AVAILABILITY

The datasets analyzed for this study can be found in https://github.com/brunobian/NeuralBasesProverbReading2018 and in http://reading.liaa.dc.uba.ar. More detailed or complementary data are available on request.

AUTHOR CONTRIBUTIONS

BB, DS, and JK designed the study and wrote the manuscript. BB collected and analyzed the data. BB and JK discussed and interpreted the results.

FUNDING

The authors were supported by the National Science and Technology Research Council (CONICET) and the University of Buenos Aires (UBA). The research was supported by the University of Buenos Aires (UBA) (20020170200259BA), The National Agency of Promotion of Science and Technology (PICT 2016-1256) and the National Scientific and Technical Research Council (PIP 11220150100787CO).

ACKNOWLEDGMENTS

We thank Marcos L. Pietto, Gustavo Juantorena, and Rodrigo Loredo for their assistance during data acquisition. We thank Rodrigo Loredo, Angel Tabullo, and Carolina Gattei for their thoughtful reading. We thank Thomas A. Gavin, Professor Emeritus, Cornell University, for helping with the edition of English in this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00082/full#supplementary-material

REFERENCES

Baayen, H. (2008). *Analyzing Linguistic Data*. Cambridge, UK: Cambridge University Press.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Statis. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boston, M. F., Hale, J., Kliegl, R., Patil, U., and Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: an evaluation using the potsdam sentence corpus. *J. Eye Move. Res.* 2, 1–12. doi: 10.16910/jemr.2.1.1

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vision* 10, 433–436. doi: 10.1163/156856897X00357

Brouwer, H. and Hoeks, J. C. (2013). A time and place for language comprehension: mapping the n400 and the p600 to a minimal cortical network. *Front. Hum. Neurosci.* 7:758. doi: 10.3389/fnhum.2013.00758

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verb. Learn. Verb. Behav.* 12, 335–359. doi: 10.1016/S0022-5371(73)80014-3

de Lange, F. P., Jensen, O., and Dehaene, S. (2010). Accumulation of evidence during sequential decision making: the importance of top–down factors. *J. Neurosci.* 30, 731–738. doi: 10.1523/JNEUROSCI.4080-09.2010

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Fernández, G., Shalom, D. E., Kliegl, R., and Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: the incoming word predictability effect. *Lang. Cogn. Neurosci.* 29, 260–273. doi: 10.1080/01690965.2012.760745

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., and Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of erp data. *Neuroimage* 30, 1383–1400. doi: 10.1016/j.neuroimage.2005.11.048

Hohenstein, S. and Kliegl, R. (2013). *remef: Remove Partial Effects. R Package Version 1.0.6.9000*. Available online at: https://github.com/hohenstein/remef/

Kamienkowski, J. E., Varatharajah, A., Sigman, M., and Ison, M. J. (2018). Parsing a mental program: Fixation-related brain signatures of unitary operations and routines in natural visual search. *NeuroImage* 183, 73–86. doi: 10.1016/j.neuroimage.2018.08.010

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception* 36, 1–16.

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (erp). *Ann. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M. and Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.

Kveraga, K., Ghuman, A. S., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168. doi: 10.1016/j.bandc.2007.06.007

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the n400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. doi: 10.1016/j.tics.2004.03.008

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Molinaro, N., Barraza, P., and Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: eeg correlates of processing expected words in sentences. *NeuroImage* 72, 120–132. doi: 10.1016/j.neuroimage.2013.01.031

Molinaro, N. and Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biol. Psychol.* 83, 176–190. doi: 10.1016/j.biopsycho.2009.12.006

Molinaro, N., Vespignani, F., Canal, P., Fonda, S., and Cacciari, C. (2008). Cloze probability does not only affect n400 amplitude: the case of complex prepositions. *Psychophysiology* 45, 1008–1012. doi: 10.1111/j.1469-8986.2008.00694.x

Monsalve, I. F., Pérez, A., and Molinaro, N. (2014). Item parameters dissociate between expectation formats: a regression analysis of time-frequency decomposed eeg data. *Front. Psychol.* 5:847. doi: 10.3389/fpsyg.2014.00847

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vision* 10, 437–442.

Pernet, C. R., Chauveau, N., Gaspar, C., and Rousselet, G. A. (2011). Limo eeg: a toolbox for hierarchical linear modeling of electroencephalographic data. *Comput. Intell. Neurosci.* 2011:831409. doi: 10.1155/2011/831409

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Smith, N. J. and Kutas, M. (2015a). Regression-based estimation of erp waveforms: I. the rerp framework. *Psychophysiology* 52, 157–168. doi: 10.1111/psyp.12317

Smith, N. J. and Kutas, M. (2015b). Regression-based estimation of erp waveforms: Ii. nonlinear effects, overlap correction, and practical considerations. *Psychophysiology* 52, 169–181. doi: 10.1111/psyp.12320

Taylor, W. L. (1953). cloze procedure : a new tool for measuring readability. *J. Bull.* 30, 415–433.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *J. Cogn. Neurosci.* 22, 1682–1700. doi: 10.1162/jocn.2009.21293

Widmann, A. and Schröger, E. (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. *Front. Psychol.* 3:233. doi: 10.3389/fpsyg.2012.00233

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 22 March 2019
doi: 10.3389/fnhum.2019.00096

Production Variability and Categorical Perception of Vowels Are Strongly Linked

Sara-Ching Chao, Damaris Ochoa and Ayoub Daliri*

Speech and Hearing Science, College of Health Solutions, Arizona State University, Tempe, AZ, United States

Theoretical models of speech production suggest that the speech motor system (SMS) uses auditory goals to determine errors in its auditory output during vowel production. This type of error calculation indicates that within-speaker production variability of a given vowel is related to the size of the vowel's auditory goal. However, emerging evidence suggests that the SMS may also take into account perceptual knowledge of vowel categories (in addition to auditory goals) to estimate errors in auditory feedback. In this study, we examined how this mechanism influences within-speaker variability in vowel production. We conducted a study (n = 40 adults), consisting of a vowel categorization task and a vowel production task. The vowel categorization task was designed—based on participant-specific vowels—to estimate the categorical perceptual boundary (CPB) between two front vowels (/ɛ/ and /æ/). Using the vowel production data of each participant, we calculated a variability-based boundary (VBB) located at the "center of mass" of the two vowels. The inverse of the standard deviation of a vowel distribution was used as the "mass" of the vowel. We found that: (a) categorical boundary was located farther from more variable vowels; and (b) the calculated VBB (i.e., the center of mass of the vowels) significantly and positively correlated with the estimated categorical boundary (r = 0.912 for formants calculated in hertz; r = 0.854 for formants calculated in bark). Overall, our findings support a view that vowel production and vowel perception are strongly and bidirectionally linked.

Keywords: speech, perception, variability, speech motor control, vowels

Edited by:
Xing Tian,
New York University Shanghai, China

Reviewed by:
Nicole Eva Neef,
Max Planck Institute for Human Cognitive and Brain Sciences, Germany
John F. Houde,
University of California, San Francisco, United States

Correspondence:
Ayoub Daliri
ayoub.daliri@asu.edu

Received: *02 November 2018*
Accepted: *01 March 2019*
Published: *22 March 2019*

Citation:
Chao S-C, Ochoa D and Daliri A (2019) Production Variability and Categorical Perception of Vowels Are Strongly Linked. Front. Hum. Neurosci. 13:96. doi: 10.3389/fnhum.2019.00096

INTRODUCTION

A large body of literature indicates that the speech production and speech perception systems interact in many intricate ways (Galantucci et al., 2006; Guenther, 2006; Tatham and Morton, 2006; Hickok, 2012; Perkell, 2012). Recent functional imaging studies have reported that motor regions—classically believed to be involved in movement production—are active during speech perception tasks (Wilson et al., 2004; Skipper et al., 2005; D'Ausilio et al., 2009; Grabski et al., 2013; Schuerman et al., 2017). Similarly, auditory regions—classically believed to be involved in speech perception—are active during speech production (Tourville et al., 2008; Hickok, 2012; Niziolek et al., 2013; Skipper et al., 2017). In fact, in a series of studies, we provided behavioral and electrophysiological evidence that the auditory system is prepared for its roles in speech monitoring several 100 ms prior to speech initiation (Daliri and Max, 2015, 2016, 2018; Merrikhi et al., 2018). Overall, the dynamic relationship between auditory and motor regions plays an important role in both speech production and speech perception (Houde and Nagarajan, 2011; Guenther and Vladusich, 2012; Hickok, 2012; Houde and Chang, 2015; Daliri et al., 2018).

Current theoretical models of speech production suggest that vowel production is strongly reliant on internally represented speech goals (Houde and Nagarajan, 2011; Hickok, 2012; Guenther, 2016). Although the exact nature of the speech goals is unclear, it has been suggested that the speech motor system (SMS) may use perceptual goals (e.g., auditory goals) to determine errors in its motor output (Perkell et al., 1997, 2008; Perkell, 2012; Guenther, 2016). These models posit that during production, the SMS compares auditory feedback of the produced speech with its auditory goals; when the auditory feedback resides outside the auditory goals (i.e., auditory error), the SMS generates corrective motor responses to reduce the perceived error. One prediction of such conceptualization is that speakers with smaller auditory goals would have smaller production variability. Because auditory goals cannot be measured directly, auditory acuity measures—estimated *via* speech discrimination tasks—have been adopted as proxies for auditory goals (Perkell et al., 2004a; Villacorta et al., 2007; Feng et al., 2011; Perkell, 2012; Daliri et al., 2013; Franken et al., 2017). In a speech discrimination task, speakers are asked to discriminate between speech sounds with subtle acoustic differences; therefore, discrimination tasks measure speakers' ability to distinguish small changes in auditory input (i.e., auditory acuity). In support of this prediction, a few studies have examined the relationship between auditory acuity and vowel production (Perkell et al., 2004a, 2008; Franken et al., 2017). These studies have shown that speakers with better auditory acuity—typically interpreted as smaller auditory goals regions—produce more consistent vowels (i.e., smaller within-vowel variability).

This type of interpretation is in line with phonetic theories that rely on local constraints (e.g., in articulatory-acoustic-perceptual space) to explain how phonological systems emerge (Stevens, 1989; Stevens and Keyser, 2010). However, it has been argued that phonological systems can also emerge based on global constraints (e.g., maximizing distance between different phonemes; Liljencrants et al., 1972) or a combination of local and global constraints (Schwartz et al., 1997, 2005). It is conceivable, therefore, to argue that the SMS may also use global constraints in addition to local constraints to more accurately produce vowels or phonemes, in general. In fact, emerging evidence suggests that the SMS may rely on perceptual knowledge of vowel categories to estimate errors in auditory feedback (Niziolek and Guenther, 2013; Bourguignon et al., 2014, 2016; Lametti et al., 2014a). For example, in a seminal study, Niziolek and Guenther (2013) showed that real-time auditory feedback perturbations (shifts in formant frequencies) of productions that were closer to the edge of the vowel category elicited larger compensatory responses relative to identical perturbations of productions closer to the center of the vowel (far from the edge of the vowel boundary). These results suggested that the SMS may use the perceptual boundary between two adjacent vowels—in addition to auditory goals—to determine errors in its output. Certainly, auditory feedback perturbations provide valuable insights into the mechanisms of error calculation in response to altered auditory feedback; however, it is not clear how this type of error calculation influences within-speaker variability in vowel production with normal, unaltered auditory feedback.

Generally, perceptual distinctiveness of two phonemes depends on the distance of each of the phonemes from their joint categorical boundary. For example, two cross-boundary tokens (e.g., /ɛ/ and /æ/) that are close to the categorical boundary between them are less distinct than two tokens that are far from the categorical boundary and close to their centroids (Kuhl, 1991; Kuhl et al., 2007; Goldstone and Hendrickson, 2010). In this study, we examined whether the categorical boundary between two adjacent vowels is related to variability of the vowels—in two adjacent vowels, the vowel closer to the categorical boundary is less variable than the vowel farther from the categorical boundary. In other words, each vowel pushes the perceptual boundary away based on the inverse of its variability. In an analogy to physics, two adjacent vowels can be considered two connected masses, and the "mass" of each vowel can be determined by the inverse of the variability of the vowel distribution. Based on this analogy, we hypothesized that variabilities of two adjacent vowels may co-vary with the categorical boundary between the vowels, and the "center of mass" of the vowel categories correlates with the categorical perceptual boundary (CPB). To test this hypothesis, we conducted a standard categorical perception task to estimate the perceptual boundary between /ɛ/ and /æ/. Given that our goal was to examine each participant's perception in relation to the participant's production variability, we used a participant-specific speech sample to generate a participant-specific vowel continuum for the categorical perception task. We also conducted a vowel production task (/ɛ/ and /æ/) and calculated the variability of each of the vowels and combined the variabilities to construct a theoretical variability-based boundary (VBB; the center of mass of vowel distributions). We found that the calculated VBB positively and strongly correlated with the perceptual boundary.

MATERIALS AND METHODS

Participants

Forty healthy adult speakers (29 female participants; M_{age} = 24.07 years, SD_{age} = 4.67 years; age range 18.42–43.01 years) participated in this study. Participants were native speakers of American English with no history of neurological, psychological, speech-language disorders, and hearing disorders (pure tone hearing threshold $\leq$20 dB HL at octave frequencies from 250 to 8,000 Hz). The Institutional Review Board at Arizona State University approved all study protocols. Participants signed a consent form prior to participation in the experiment. Participants were recruited from a participant pool of undergraduate students.

Procedure

Participants were seated inside a sound booth in front of a computer monitor. A microphone (SM58, Shure) mounted on a stand was placed 15 cm from the corner of the participant's mouth (at ∼45° angle). The microphone signal was amplified (Tubeopto 8, ART) and digitized (at 48,000 Hz sampling rate) *via* an audio interface (Ultralite Mk3 hybrid, MOTU). Output signals of the audio interface were then amplified (Pro Rx1602, Eurorack) and played back to the participant *via* insert earphones (ER-1, Etymotic Research Inc.). The input-output level was calibrated prior to each experiment to ensure that the intensity of the played-back signal was 5 dB greater than the microphone signal.

Each participant completed the study in one session that took less than 30 min. Participants completed a practice task in which they overtly produced monosyllabic consonant-vowel-consonant (CVC) words (e.g., "head"). The practice task (30 trials) was used to familiarize participants with the setup and to train them to pronounce target words within a desired intensity (70–80 dB SPL) and duration (400–600 ms; based on the voiced segments) range. After each trial, participants received visual feedback regarding their intensity and duration. Next, participants completed a *vowel production task* that was similar to the practice task. Participants produced CVC words that contained /ε/ or /æ/ (30 trials of each vowel). The order of words (vowels) was randomized. In this task, if participants produced words within the desired intensity and duration ranges, they did not receive visual feedback.

Upon completion of the vowel production task, for each produced word in the vowel production task, we extracted the first formant frequency (F1) and the second formant frequency (F2) from vowels of each word. We used Audapter—a publicly available software for formant tracking and manipulation—to automatically extract the formant frequencies (Cai, 2015). Audapter is a MATLAB-based software package that its source code is implemented in C++ and consists of several speech processing blocks, including formant tracking and formant manipulation. Audapter uses linear predictive coding (LPC) analysis and dynamic programming to track formant frequencies. We used LPC order of 17 for male participants and 15 for female participants. The speech data was recorded at 48,000 Hz and down-sampled to 16,000 Hz to reduce computational loads. To improve formant-tracking accuracy, we supplied Audapter with participant-specific initial values for F1 and F2 (in Hz) that were estimated based on the practice trials. Audapter uses smoothed short-term energy criteria in combination with heuristic rules to determine onset and offset of voiced segments and to initiate formant tracking and formant manipulations. After the vowel production task, a custom written algorithm used onset and offset values determined by Audapter and extracted the average formant values (in Hz) in a window placed on the center of the segment (10%–90% into the length of the segment). Using F1-F2 coordinates, the algorithm used the Euclidian distance to determine the token closest to the median of the vowel /ε/ and the median of the vowel /æ/ (hereafter called median productions). In other words, *median productions* of a given participant were words produced by the participant that were closest to the center of the distribution of the vowel /ε/ and the center of distribution of the vowel /æ/ of the participant (in F1-F2 coordinates). Using these participant-specific median productions, we generated a set of six or seven equally spaced stimuli (formant shifted CVC words) along the line connecting the median /ε/ and the median /æ/ for each participant. Given that samples were generated based on participant specific speech, the duration of stimuli were different for different participants. The duration of the voiced segment of stimuli ranged from ∼382 ms to ∼627 ms (M = 472 ms, SD = 39 ms). However, for a given participant, only the vowel portions of the stimuli (words) were different, as the stimuli were generated based on the participant-specific median production by shifting F1 and F2 of the median /ε/ (using offline formant shift of Audapter). The stimuli were designed such that the vowel of the first stimulus coincided with the median /ε/ and the vowel of the last stimulus coincided with the median /æ/. **Figure 1A** shows a set of six stimuli for a representative participant that are distributed along the line connecting the two vowels. We then used these participant-specific speech stimuli in a standard *categorical perception task* (Möttönen and Watkins, 2009; Niziolek and Guenther, 2013). Each stimulus was presented 10 times and the order of stimuli was randomized. In each trial of the perception task, a token from the participant-specific stimuli set was presented to the participant (at 75 dB SPL) and he/she was asked to indicate (using a keypad) which word was presented (e.g., "head" or "had"; ε or æ).

Data Analysis

The goal of this study was to examine the relationship between vowel perception and variability of vowel production. We used data from the categorical perception task and the vowel production task to drive participant-specific perception-based boundary and production-based boundary between /ε/ and /æ/.

Categorical Perceptual Boundary (CPB)

We fitted a logistic psychometric function to each participant's response (proportion of /ε/ responses) using a Maximum Likelihood criterion (Kingdom and Prins, 2016; Prins and Kingdom, 2018). Evaluation of the goodness of fitted psychometric functions confirmed that the psychometric functions were fitted properly for all participants ($R^2 > 0.85$). Based on the fitted psychometric functions, we then estimated each participant's CPB—formant values at 50% proportion /ε/-responses. We used six levels of stimuli for 27 participants and seven levels stimuli for 13 participants; our analyses did not reveal a statistically significant difference between extracted CPBs of these two groups (p = 0.45), and thus, we combined both groups for further analyses. **Figure 1B** shows the fitted psychometric functions of all participants, along with their estimated perceptual boundaries (shown as circles).

Variability-Based Boundary (VBB) or Center of Mass

To examine variability of the two vowels, we implemented the following steps. Note that these analyses were done

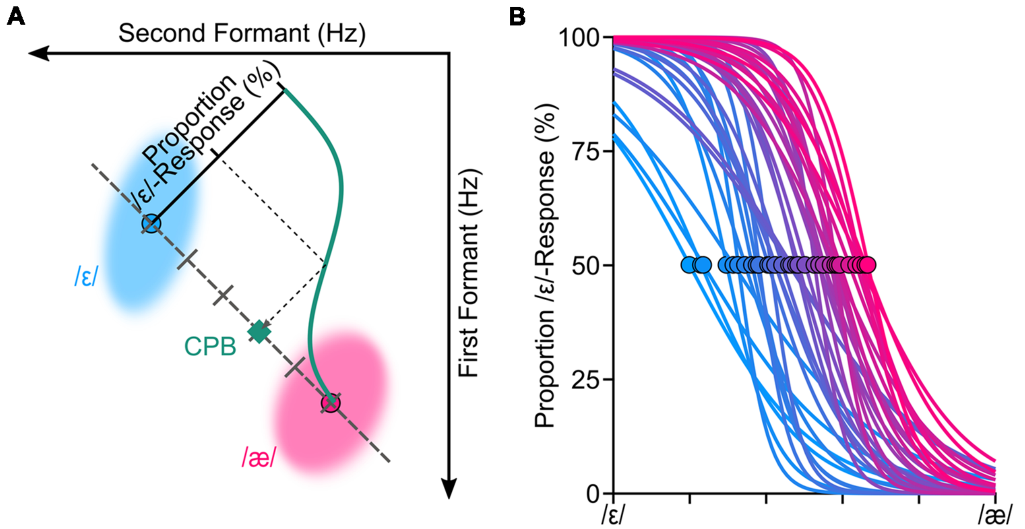


FIGURE 1 | We conducted a standard categorical perceptual task to estimate the categorical perceptual boundary (CPB) between /ɛ/ and /æ/ for each participant. Panel **(A)** shows the procedure for estimating the CPB (at 50%) through fitting a logistic psychometric function to the perceptual data for one representative participant. To generate speech stimuli for the categorical perception task, we manipulated participants' own production samples (formant shifted) along the line connecting participant-specific centroids of /ɛ/ and /æ/. The shaded areas (ellipses) in **(A)** correspond to the spread of the formants in hertz (two standard deviations from the mean of principle components) of /ɛ/ and /æ/ for this representative participant. Psychometric functions were successfully fitted ($R^2 > 0.85$) to all participants' responses **(B)**. The circles in **(B)** correspond to where CPBs are located for all participants.

offline upon completion of the study, and they are different from the initial formant analysis that was done during the experimental session. First, all productions were inspected (offline) to exclude gross errors in formant tracking and to exclude trials with speech errors (e.g., producing wrong words). Approximately, 1% of all trials were excluded. Second, based on the spectrogram of each production, onset and offset of vowels were manually annotated and F1 and F2 trajectories were extracted. To extract formants, we averaged formant values from a window placed on the center of the vowel (40%–60% into the length of the vowel; steady-state portion of the vowel). Third, we projected F1 and F2 values of each produced vowel to a line connecting participant-specific median /ɛ/ to median /æ/. We used median /ɛ/ as a reference point for all projected formant values. The rationale for this procedure was to estimate the variability of the vowels along the line connecting the two vowels, as the stimuli set used in the perception task was generated along this line. Thus, this procedure ensured that vowel variability and perceptual results were along the same line and based on participant-specific vowel configurations. Fourth, given that we hypothesized that the CPB between the two vowels may co-vary with vowel variability, we used vowel variabilities to estimate a VBB. **Figure 2A** shows the procedure for the calculation of the VBB for one representative participant. The VBB was defined as the center of mass between the two vowel distributions, and the mass of each vowel was the inverse of its variability (standard deviation along the line connecting the two vowels). In other words, the VBB is a theoretical boundary between two vowel distributions and was calculated based on variabilities of the distributions. In these calculations, the VBB was calculated relative to the center of /ɛ/ (reference point).

$$VBB = \frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon} + \sigma_{æ}} D_{\varepsilon-æ}$$

In this formula, σ represents the standard deviation of the vowels and $D_{\varepsilon-æ}$ represents the distance between the vowel centroids. It should be noted that another approach to arrive to the same equation is based on the normalized distance of the VBB from each of the distributions. The VBB is the point between two vowels where its distance from /ɛ/ distribution is the same as its distance from /æ/ distribution.

$$D_{vbb-\varepsilon} = \frac{|VBB - \mu_{\varepsilon}|}{\sigma_{\varepsilon}}, D_{vbb-æ} = \frac{|VBB - \mu_{æ}|}{\sigma_{æ}}$$

In these equations, μ corresponds to the mean of a vowel distribution. Given that we calculated projected formants relative to /ɛ/, then $\mu_{\varepsilon} = 0, \mu_{æ} = D_{\varepsilon-æ}$, and VBB $< \mu_{æ}$; thus, we can simplify the equations and arrive at the equation for the center of mass.

$$D_{vbb-\varepsilon} = \frac{VBB}{\sigma_{\varepsilon}}, D_{vbb-æ} = \frac{D_{\varepsilon-æ} - VBB}{\sigma_{æ}}, D_{vbb-æ} = D_{vbb-\varepsilon}$$

During the study, all formants were measured in hertz and speech stimuli for the perception task were calculated in hertz. However, to ensure that the relationship between the perception and production measures were valid in psychoacoustic scales, we transformed formant values from hertz to bark (Traunmüller, 1990) and followed the same steps to estimate the VBB in bark scale. We also used a similar projection procedure to calculate the CPB in hertz and in bark for each participant.

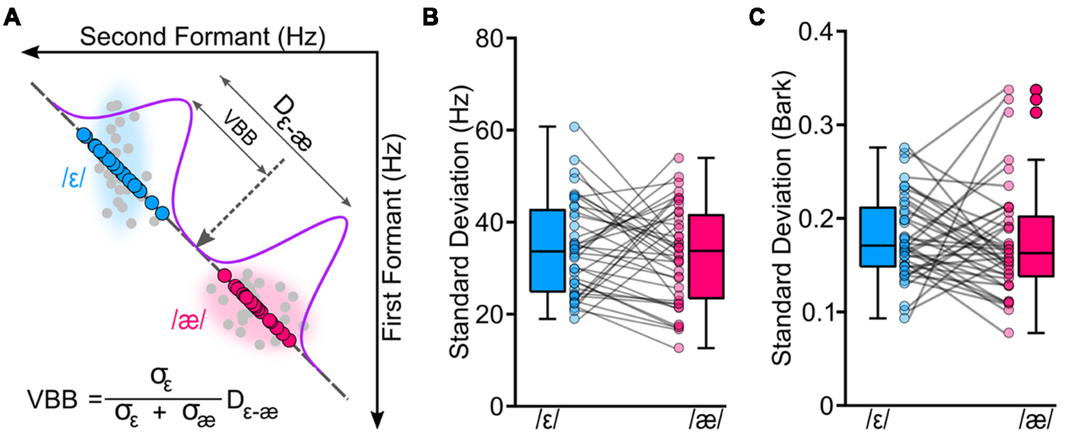


FIGURE 2 | We conducted a vowel production task to estimate a theoretical variability-based boundary (VBB or the center of mass of the two vowels) between /ε/ and /æ/ for each participant. Panel **(A)** shows the original formant values (in hertz) of the two vowels for a representative participant (gray circles). To calculate VBB, we projected formant values of each participant to a line connecting the two vowels; blue circles correspond to projected formant values of /ε/, and magenta circles correspond to projected formant values of /æ/. Then, we estimated standard deviations of the vowels (σ_ε and $\sigma_{æ}$) along with the distance between centroids of the two vowels ($D_{\varepsilon\text{–}æ}$). Our analyses showed that two vowels had similar variabilities for formants measured in hertz (**B**; $p = 0.507$) and in bark (**C**; $p = 0.694$).

The estimated CPB and VBB values were entered in statistical analyses. Note that this study was not designed to examine whether or not perception drives production, and the association between the two systems was treated from a correlational perspective. Prior to analyses, we performed the Shapiro-Wilk test to ensure normality of all data. We used Pearson's correlation coefficients and regression analyses to examine relationship between the VBB and the CPB. We examined residual values to confirm linear model assumptions. Additionally, we used paired *t*-tests to compare vowel variabilities. R version 3.5.1 (The R Project for Statistical Computing[1]) was used for all statistical analyses.

RESULTS

Table 1 shows the group average F1, F2, and projected formants in both hertz and bark. Most relevant to our analyses are the projected formant values. Note that these projected formant values were calculated relative to the center (median) of /ε/ for each participant. The average ε–æ distance (Euclidean distance in F1-F2 coordinates) in hertz was 249.09 Hz (*SD* = 80.50; 128.61–434.61), and in bark was 1.38 bark (*SD* = 0.37; 0.81–2.35). As shown in **Figures 2B,C**, we did not find a statistically significant difference between the variability of /ε/ and the variability of /æ/ (standard deviation of projected formants; σ_ε and $\sigma_{æ}$) in hertz ($t_{(39)} = 0.670$, $p = 0.507$) and in bark ($t_{(39)} = 0.396$, $p = 0.694$). Estimated CPB in hertz ranged from 34.16 to 234.51 Hz (*M* = 121.62 Hz, *SD* = 52.80) and in bark ranged from 0.21 to 1.33 bark (*M* = 0.69, *SD* = 0.27). The calculated VBB in hertz ranged from 44.81 to 256.48 Hz (*M* = 128.29, *SD* = 48.73) and in bark ranged from 0.35 to 1.33 bark (*M* = 0.71 bark, *SD* = 0.25). No statistically significant difference was found between the CPB and the VBB in hertz ($t_{(39)} = 1.945$, $p = 0.061$) or in bark ($t_{(39)} = 1.022$, $p = 0.313$).

The primary goal of this study was to examine relationships between perceptual boundary (measured in categorical perception task) and VBB (calculated based on data from the vowel production task). First, to test whether the CPB is farther from more variable vowels, we examined the relationship between standard deviation of /ε/ normalized by the sum of the standard deviations of /ε/ and /æ/ [i.e., $\sigma_\varepsilon/(\sigma_\varepsilon + \sigma_{æ})$], using Pearson's correlation coefficients (*r*). We found statistically significant positive correlation between the CPB and the normalized vowel variability (for data in hertz: $r = 0.425$, $p = 0.006$; for data in bark: $r = 0.426$, $p = 0.006$). This result suggested that the CPB was closer to the center of /ε/ in speakers with less variable /ε/ (relative to /æ/), and farther from the center of /ε/ in speakers with more variable /ε/. Second, as mentioned in the method section, the VBB is a theoretical boundary located between two vowels such that its distance from /ε/ distribution is the same as its distance from /æ/ distribution. In other words, the VBB is the optimal point between the two vowel distributions (i.e., the center of mass of the two distributions). Using Pearson's correlation coefficients, we found that the CPB strongly and positively correlated with the VBB both in hertz ($r = 0.912$, $p < 0.001$) and in bark ($r = 0.854$, $p < 0.001$). As shown in **Figure 3**, the two methods of calculation (hertz and bark) resulted in similar outcomes. Third, we used regression analyses to: (a) estimate the slope value; and (b) examine how much of the variability of the CPB can be explained by the VBB. We conducted a simple linear regression to test whether the VBB in hertz predicted the CPB in hertz (CPB = Slope × VBB + Intercept). We found that the VBB in hertz explained 82.7% of the variance of the CPB in hertz ($R^2 = 0.827$, $F_{(1,38)} = 187.3$, $p < 0.001$), with a statistically significant slope value of 0.988 ($p < 0.001$). Similarly, we found that the VBB in bark explained 72.2% of the variance of the CPB

---

[1] www.R-project.org

TABLE 1 | Group average and standard deviation (inside parentheses) of the formant values and projected formant values for /ɛ/ and /æ/ in hertz and bark.

	F1		F2		Projected formants	
	Hz	**Bark**	**Hz**	**Bark**	**Hz**	**Bark**
/ɛ/	721.87 (100.66)	6.53 (0.78)	1904.99 (153.75)	12.76 (0.55)	0.15 (1.81)	0.00 (0.01)
/æ/	902.79 (131.97)	7.83 (0.91)	1731.40 (152.78)	12.13 (0.62)	249.21 (80.55)	1.39 (0.37)

The center of /ɛ/ was used as the reference point for projected formant values.

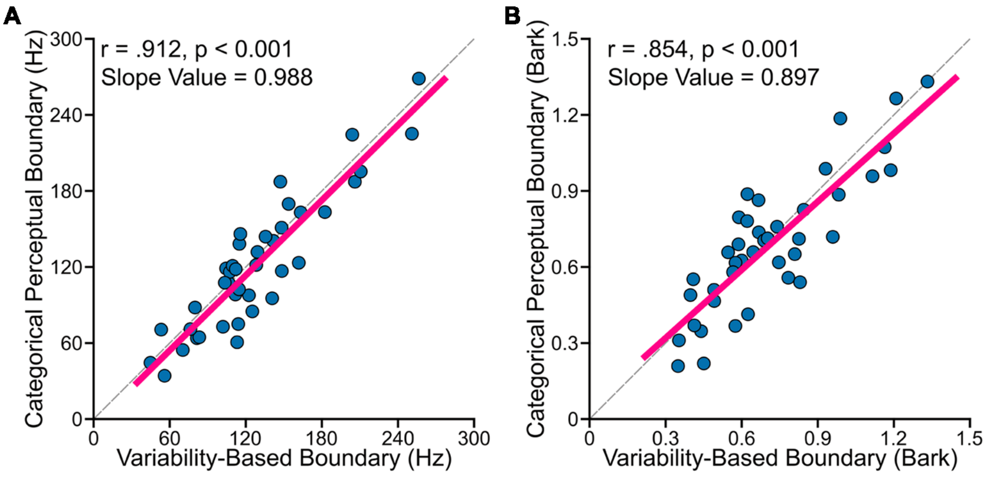


FIGURE 3 | Individual participants' data showed (**A**: formants calculated in hertz; **B**: formants calculated in bark) that VBB (or the center of mass of two vowels) strongly and positively correlated with the CPB for data measured in hertz (*r* = 0.912, *p* < 0.001) and in bark (*r* = 0.854, *p* < 0.001). The dashed lines in **(A,B)** correspond to lines with the slope of one.

in bark ($R^2 = 0.722$, $F_{(1,38)} = 102.3$, $p < 0.001$), with a statistically significant slope value of 0.897 ($p < 0.001$).

DISCUSSION

Previous studies have provided behavioral and neural evidence for the link between the speech production and speech perception systems (Galantucci et al., 2006; Tatham and Morton, 2006; Hickok, 2012; Perkell, 2012; Guenther, 2016). In the present study, we examined whether the perception of two adjacent vowels interacts with the production variabilities of the two vowels. We conducted a standard categorical perception task to estimate the CPB between /ɛ/ and /æ/ using participant-specific speech samples. We also conducted a vowel production task to determine participant-specific variabilities of /ɛ/ and /æ/. In an analogy to physics, two adjacent vowels can be considered as two connected masses, and the "mass" of each vowel can be determined by the inverse of the variability of the vowel distribution. Based on this analogy, we hypothesized variabilities of two adjacent vowels may co-vary with the categorical boundary between the vowels, and the "center of mass" of the vowel categories (VBB) correlates with the CPB. Consistent with our hypotheses, we found that: (a) the CPB was farther from more variable vowels and closer to less variable vowels; and (b) the CPB strongly correlated with the VBB, and that the VBB explained 72%–82% of the variance of the CPB.

One interpretation of these results is that the SMS uses the CPB between two adjacent vowels—in addition to auditory goals—to determine errors in its auditory output, and thus, to constrain vowel variabilities. An alternative interpretation is that our productions shape our perception, and thus, vowel production variability drives categorical perception between adjacent vowels. It should be noted that these two interpretations are not mutually exclusive. This study was not designed to examine whether or not perception drives production, and the association between the two systems was treated from a correlational perspective; therefore, our results cannot rule out any of these interpretations. Empirical results and theoretical frameworks have shown that we acquire our auditory goals during infancy and childhood, and then, we use the acquired auditory goals to drive the speech production system (Callan et al., 2000; Kuhl, 2004; Guenther and Vladusich, 2012; Guenther, 2016). However, this mechanism may be different during adulthood. Based on our findings and previous reports of the close association between the perception and production systems (Tatham and Morton, 2006; Guenther, 2016), we propose that the link between speech perception and speech production is dynamic and the two systems bi-directionally influence each other. In this view, after the speech acquisition stage, the perception system and the production system seamlessly "converge" together. Therefore, change in one system could result in change in the other system—although the required magnitude and duration of exposure to a change in one system to result in a similar change in the other system is not necessarily equal for the two systems. One outcome of the convergence of the two systems is that the categorical perceptual boundary and production variability change until they reach an equilibrium at which the categorical boundary is located at the most optimal point between the two vowels. The position of this optimal point is related to both the Euclidian distance and the variability of vowels (defined as the VBB or the center of mass in this study).

Our results are largely in agreement with previous studies that have examined the relationship between speech production and perception (Newman, 2003; Perkell et al., 2004b, 2008; Nieto-Castanon et al., 2005; Franken et al., 2017). Such studies have typically used discrimination tasks to find perceptual acuity of a given vowel. The rationale for using discrimination tasks is primarily based on theoretical frameworks of speech production (Perkell et al., 1997; Guenther and Vladusich, 2012; Perkell, 2012; Guenther, 2016). For example, the Directions Into Velocities of Articulators (DIVAs) model of speech production suggests that speech units are partially represented as auditory goals, and that the auditory feedback during speech production is compared to the auditory goals (Guenther, 2016). If there is a discrepancy between the auditory goals and the incoming auditory feedback, then the brain issues a corrective motor command to compensate for the error. Based on this account of speech production, speakers with smaller auditory goals would be more sensitive to errors, which could lead to more precise and consistent speech production (i.e., less variable speech). Thus, this interpretation implies that variability of a given vowel is solely related to auditory goals of the vowel, and characteristics of adjacent vowels may not affect the vowel variability. However, there is emerging evidence (Mitsuya et al., 2011; Niziolek and Guenther, 2013; Bourguignon et al., 2014, 2016; Lametti et al., 2014b; Reilly and Pettibone, 2017) that speakers are more sensitive to experimentally induced auditory errors (through formant perturbations) that are more similar to adjacent vowels, suggesting that the SMS may also calculate "categorical errors"—i.e., whether or not the received auditory feedback of a vowel is within the vowel's perceptual category. If this is the case, then the CPB between vowels may also serve as a boundary (or a constraint) for vowel variability (i.e., productions can be variable as long as they are within the perceptual category of the vowel). In other words, for two adjacent vowels to remain perceptually distinct, if one vowel becomes more variable, then the adjacent vowel needs to become less variable to keep the two vowels distinct. Overall, our finding of a strong relationship between the VBB and the CPB supports the view that the SMS may also use the CPB (in addition to auditory acuity) to calculate auditory errors which in turn determines/limits variabilities of adjacent vowels. This interpretation does not imply that the interaction of the perception and production is unidirectional; in fact, as we mentioned above, perception and production could influence each other dynamically and bidirectionally throughout life.

Of course, our procedure and analyses have several important limitations that require further research. First, we generated speech stimuli based on participants' own median productions of /ɛ/ and /æ/, but the stimuli were generated in the F1-F2 coordinates and higher formants were not modified (e.g., F3 and F4). This may have influenced the quality of the stimuli and added some unwanted variability in the calculation of the CPB. Second, we limited our calculation to variabilities along the line connecting the two vowels (ɛ-æ line) for simplicity purposes; however, different vowels have different distributions and this simplification may have influenced the relationship between production variability and the CPB. Third, our study was designed to examine only two vowels, and it is unclear if this effect can be observed in other vowels. Future studies can overcome such limitations by: (a) manipulating all formants (and not just F1 and F2) to generate more accurate speech stimuli; (b) calculating the CPB and vowel variabilities along different pathways between vowels to estimate the relationship in the entire multi-dimensional formant space; and (c) examining all vowels in the English language as well as vowels in other languages.

In sum, we conducted a categorical perception task and a vowel production task to examine whether vowel perception correlates with vowel production variability. We found that the categorical boundary was farther from more variable vowels and closer to less variable vowels. Additionally, we found that the center of mass of two vowels (a theoretical boundary calculated based on production variability) strongly and positively correlated with the categorical boundary and it explained 72%–82% of the variance of the categorical boundary. Overall, our findings support a view that the speech perception and speech production systems are strongly and bidirectionally linked.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AD designed the experiment. S-CC, DO, and AD conducted the experiment, interpreted the results and wrote the manuscript. S-CC and AD analyzed the data.

FUNDING

This work was supported by Arizona State University New Faculty Startup Funding to AD.

REFERENCES

Bourguignon, N. J., Baum, S. R., and Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Front. Hum. Neurosci.* 8:208. doi: 10.3389/fnhum.2014.00208

Bourguignon, N. J., Baum, S. R., and Shiller, D. M. (2016). Please say what this word is-vowel-extrinsic normalization in the sensorimotor control of speech. *J. Exp. Psychol. Hum. Percept. Perform* 42, 1039–1047. doi: 10.1037/xhp0000209

Cai, S. (2015). Audapter. Available online at: https://github.com/shanqing-cai/audapter_matlab.

Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J. Speech Lang. Hear. Res.* 43, 721–736. doi: 10.1044/jslhr.4303.721

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Daliri, A., and Max, L. (2015). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain Lang.* 143, 59–68. doi: 10.1016/j.bandl.2015.03.002

Daliri, A., and Max, L. (2016). Modulation of auditory responses to speech vs. nonspeech stimuli during speech movement planning. *Front. Hum. Neurosci.* 10:234. doi: 10.3389/fnhum.2016.00234

Daliri, A., and Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex* 99, 55–68. doi: 10.1016/j.cortex.2017.10.019

Daliri, A., Prokopenko, R. A., and Max, L. (2013). Afferent and efferent aspects of mandibular sensorimotor control in adults who stutter. *J. Speech Lang. Hear. Res.* 56, 1774–1778. doi: 10.1044/1092-4388(2013/12-0134)

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., and Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Dev. Sci.* 21:e12521. doi: 10.1111/desc.12521

Feng, Y., Gracco, V. L., and Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *J. Neurophysiol.* 106, 667–679. doi: 10.1152/jn.00638.2010

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., and Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor control. *J. Acoust. Soc. Am.* 142:2007. doi: 10.1121/1.5006899

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/bf03193857

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 69–78. doi: 10.1002/wcs.26

Grabski, K., Schwartz, J., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., et al. (2013). Shared and distinct neural correlates of vowel perception and production. *J. Neurolinguistics* 26, 384–408. doi: 10.1016/j.jneuroling.2012.11.003

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* 39, 350–365. doi: 10.1016/j.jcomdis.2006.06.013

Guenther, F. H. (2016). *Neural Control of Speech*. Cambridge, MA: MIT Press.

Guenther, F. H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. *J. Neurolinguistics* 25, 408–422. doi: 10.1016/j.jneuroling.2009.08.006

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 13, 135–145. doi: 10.1038/nrn3158

Houde, J. F., and Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* 33, 174–181. doi: 10.1016/j.conb.2015.04.006

Houde, J. F., and Nagarajan, S. S. (2011). Speech production as state feedback control. *Front. Hum. Neurosci.* 5:82. doi: 10.3389/fnhum.2011.00082

Kingdom, F. A. A., and Prins, N. (2016). *Psychophysics: A Practical Introduction* 2nd edition. London: Elsevier Academic Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2007). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Lametti, D. R., Krol, S. A., Shiller, D. M., and Ostry, D. J. (2014a). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychol. Sci.* 25, 1325–1336. doi: 10.1177/0956797614529978

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., and Ostry, D. J. (2014b). Plasticity in the human speech motor system drives changes in speech perception. *J. Neurosci.* 34, 10339–10346. doi: 10.1523/jneurosci.0108-14.2014

Liljencrants, J., Lindblom, B., and Language, B. L. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48, 839–862. doi: 10.2307/411991

Merrikhi, Y., Ebrahimpour, R., and Daliri, A. (2018). Perceptual manifestations of auditory modulation during speech planning. *Exp. Brain Res.* 236, 1963–1969. doi: 10.1007/s00221-018-5278-3

Mitsuya, T., MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *J. Acoust. Soc. Am.* 130, 2978–2986. doi: 10.1121/1.3643826

Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi: 10.1523/jneurosci.6018-08.2009

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: a preliminary report. *J. Acoust. Soc. Am.* 113, 2850–2860. doi: 10.1121/1.1567280

Nieto-Castanon, A., Guenther, F. H., Perkell, J. S., and Curtin, H. D. (2005). A modeling investigation of articulatory variability and acoustic stability during American English vertical bar r vertical bar production. *J. Acoust. Soc. Am.* 117, 3196–3212. doi: 10.1121/1.1893271

Niziolek, C. A., and Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *J. Neurosci.* 33, 12090–12098. doi: 10.1523/jneurosci.1008-13.2013

Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *J. Neurosci.* 33, 16110–16116. doi: 10.1523/jneurosci.2137-13.2013

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neurolinguistics* 25, 382–407. doi: 10.1016/j.jneuroling.2010.02.011

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M. K., et al. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *J. Acoust. Soc. Am.* 116, 2338–2344. doi: 10.1121/1.1787524

Perkell, J. S., Matthies, M. L., Tiede, M. K., Lane, H., Zandipour, M., Marrone, N., et al. (2004b). The distinctness of speakers' /s/−/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *J. Speech Lang. Hear. Res.* 47, 1259–1269. doi: 10.1044/1092-4388(2004/095)

Perkell, J. S., Lane, H., Ghosh, S. S., Matthies, M. L., Tiede, M. K., Guenther, F. H., et al. (2008). "Mechanisms of vowel production: auditory goals and speaker acuity," in *Proceedings of the 8th International Seminar on Speech Production*, eds R. Sock, S. Fuchs, and Y. Laprie. (Strasbourg, France), 29–32.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., et al. (1997). Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech Commun.* 22, 227–250. doi: 10.1016/s0167-6393(97)00026-5

Prins, N., and Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Front. Psychol.* 9, 1250. doi: 10.3389/fpsyg.2018.01250

Reilly, K. J., and Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *J. Neurophysiol.* 118, 2925–2934. doi: 10.1152/jn.00702.2016

Schuerman, W. L., Meyer, A. S., and McQueen, J. M. (2017). Mapping the speech code: cortical responses linking the perception and production of vowels. *Front. Hum. Neurosci.* 11:161. doi: 10.3389/fnhum.2017.00161

Schwartz, J., Abry, C., Boe, L., Lucie, M., and Vall, N. (2005). Asymmetries in vowel perception , in the context of the dispersion - focalisation theory. *Speech Commun.* 45, 425–434. doi: 10.1016/j.specom.2004.12.001

Schwartz, J., Boë, L., Vallée, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *J. Phon.* 25, 255–286. doi: 10.1006/jpho.1997.0043

Skipper, J. I., Devlin, J. T., and Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* 164, 77–105. doi: 10.1016/j.bandl.2016.10.004

Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage* 25, 76–89. doi: 10.1016/j.neuroimage.2004.11.006

Stevens, K. N. (1989). On the quantal nature of speech. *J. Phon.* 17, 3–45.

Stevens, K. N., and Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *J. Phon.* 38, 10–19. doi: 10.1016/j.wocn.2008.10.004

Tatham, M., and Morton, K. (2006). *Speech Production and Perception*. London: Palgrave Macmillan.

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97–100. doi: 10.1121/1.399849

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J. Acoust. Soc. Am.* 122, 2306–2319. doi: 10.1121/1.2773966

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 21 May 2019
doi: 10.3389/fnhum.2019.00154

Language Brain Representation in Bilinguals With Different Age of Appropriation and Proficiency of the Second Language: A Meta-Analysis of Functional Imaging Studies

Elisa Cargnelutti[1]*, *Barbara Tomasino*[1] *and Franco Fabbro*[2,3]

[1]Scientific Institute, IRCCS E. Medea, Dipartimento/Unità Operativa Pasian di Prato, Udine, Italy, [2]Cognitive Neuroscience Laboratory, DILL, University of Udine, Udine, Italy, [3]PERCRO Perceptual Robotics Laboratory, Scuola Superiore Sant'Anna, Pisa, Italy

Language representation in the bilingual brain is the result of many factors, of which age of appropriation (AoA) and proficiency of the second language (L2) are probably the most studied. Many studies indeed compare early and late bilinguals, although it is not yet clear what the role of the so-called critical period in L2 appropriation is. In this study, we carried out coordinate-based meta-analyses to address this issue and to inspect the role of proficiency in addition to that of AoA. After the preliminary inspection of the early (also very early) and late bilinguals' language networks, we explored the specific activations associated with each language and compared them within and between the groups. Results confirmed that the L2 language brain representation was wider than that associated with L1. This was observed regardless of AoA, although differences were more relevant in the late bilinguals' group. In particular, L2 entailed a greater enrollment of the brain areas devoted to the executive functions, and this was also observed in proficient bilinguals. The early bilinguals displayed many activation clusters as well, which also included the areas involved in cognitive control. Interestingly, these regions activated even in L1 of both early and late bilingual groups, although less consistently. Overall, these findings suggest that bilinguals in general are constantly subjected to cognitive effort to monitor and regulate the language use, although early AoA and high proficiency are likely to reduce this.

Keywords: meta-analysis, bilingualism, age of appropriation (AoA), proficiency, first language (L1), second language (L2)

Edited by:
Peter Sörös,
University of Oldenburg, Germany

Reviewed by:
Mireia Hernandez,
University of Barcelona, Spain
Ekaterina Kuzmina,
University of Oslo, Norway

****Correspondence:***
Elisa Cargnelutti
elisa.cargnelutti@lanostrafamiglia.it

Received: *28 December 2018*
Accepted: *23 April 2019*
Published: *21 May 2019*

Citation:
Cargnelutti E, Tomasino B and Fabbro F (2019) Language Brain Representation in Bilinguals With Different Age of Appropriation and Proficiency of the Second Language: A Meta-Analysis of Functional Imaging Studies. Front. Hum. Neurosci. 13:154. doi: 10.3389/fnhum.2019.00154

INTRODUCTION

How does the brain of a bilingual person work? Research on this topic has extensively developed in the last decades, with an increasing number of studies devoted to the identification of the brain areas activated when bilinguals perform language tasks in the known languages. Bilinguals are different from each other in several respects. Given the definition of a bilingual as a person that masters more than one language or dialect (see Fabbro, 1999, 2001), it appears clear that this may apply to a wide range of individuals. For this reason, the bilinguals assessed in the published studies rarely form consistent and homogeneous groups. It follows that a comprehensive and universally accepted picture about the bilingual brain functioning is still lacking and that some questions are yet to be answered.

In this study, we tried to address some of these issues. The first notes the possibly different language brain representation in relation to the so defined *critical period*. This represents a specific age after which the learning process becomes challenging and the achieved performance in the second language (L2) hardly equals that of the native or first language (L1). In the past decades, an extensive debate has concerned the identification of the L2 *age of acquisition* or *age of appropriation* (as lately better defined, see Paradis, 2009) cutoff (hereafter, AoA). Some authors set it around puberty, a period during which language skills fully develop (e.g., Lenneberg, 1967; Long, 1990; Locke and Bogin, 2006), whereas others suggest the period around 6–7 years of age to be crucial, because, after this age, learning some linguistic skills becomes challenging (e.g., Johnson and Newport, 1989).

As long as AoA is judged as one of the parameters that mainly determine the L2 performance and shapes its brain representation, many studies compared the language networks between *early* and *late* bilinguals, hence between bilinguals having approached L2 either before or after the defined AoA cutoff. Although not univocally, most of these studies used 6 years of age as the AoA cutoff. This choice is motivated by the important developmental events taking place around this age. First, the brain is almost at its adult size (e.g., Giedd et al., 1999; Casey et al., 2000) and most of the myelination processes are complete (e.g., Nakagawa et al., 1998). Concerning language, skill achievement is attained in almost every domain, despite the fact that not all of the skills are perfectly mastered yet (e.g., Skeide and Friederici, 2016). Another important change concerns the memory systems supporting the cognitive processes. At this age, memory is organized as in adults and the verbal component takes on importance with respect to the visuo-spatial components (e.g., Gathercole et al., 2004).

Important changes also take place concerning the dissociation between implicit and explicit memory systems (see Paradis, 1994, 2004, 2009; Ullman, 2001, 2005, 2006). Up to this age, in fact, children acquire skills through implicit memory, therefore in an almost unconscious way. These skills are easily internalized and automatically applied. Along with development, this memory system becomes less flexible and late-learned skills are therefore mainly supported by explicit memory, with the enrollment of conscious brain processes. These skills are unlikely to become highly automatized, in particular concerning some language domains, such as grammar and phonology/articulation, whereas lexico-semantics appeared to be less affected by AoA (see also Ruben, 1997). In this case, the critical AoA seems to fall on adolescence, as long as the lexical knowledge mainly depends on the declarative memory capacity, then on proficiency and extent of use.

Another tricky aspect concerns the effects of the acquisition of both languages roughly simultaneously since birth. Such bilinguals are referred to as *simultaneous*, in comparison with the *sequential* bilinguals, who, irrespective of their AoA, had approached L2 successively to L1 and possibly when L1 acquisition was almost complete. As the majority of bilinguals belongs to the second category, neuroimaging data on the language brain networks in simultaneous bilinguals is reported in very few papers. Rather, studies more often include bilinguals having learned the two languages at least in the very first years of life (see **Supplementary Table S1**). In this sense, it would be interesting to inspect whether the language brain networks of *very early* bilinguals differ from those of general early bilinguals, as inspected in a few studies (see **Supplementary Table S1**).

A few previous meta-analyses focused on the functional networks associated with each language in the groups of early and late bilinguals. In this respect, Liu and Cao (2016) found that L2 activated several regions (i.e., insula and frontal cortex areas) more than L1 and this especially occurred in the group of late bilinguals. Similarly, Indefrey (2006), who conducted an explorative investigation of the areas that activate in bilinguals with different AoA, observed that it was more likely for individuals with late AoA to have an overall greater activation (especially in the left inferior frontal gyrus). The author reported a similar trend for bilinguals with low proficiency/exposure.

The aspect of language proficiency has been addressed in another meta-analysis (Sebastian et al., 2011). In this case, the authors observed that the L1 and L2 networks were more similar to each other in the group of the high-proficiency bilinguals, whereas greater differences between the two languages emerged as the result of low proficiency. Actually, evidence from the clinical literature seems to suggest that factors such as language proficiency and use/exposure are sometimes more relevant than mere AoA. There were indeed cases of bilingual aphasia in which the language that was premorbidly "weaker" was the most affected, whereas the language that the patient mastered better was less impaired. This indicated that proficiency in a given language is sometimes more relevant in predicting the impairment profile in bilingual aphasia (e.g., Edmonds and Kiran, 2006; Druks and Weekes, 2013; Gray and Kiran, 2013).

Nevertheless, a recently published systematic review on bilingual aphasia reported on the role of proficiency and use to be secondary to that of AoA (Kuzmina et al., 2019). Actually, L2 was more preserved—probably because of its stronger brain representation—in the case it was the best-mastered or mostly used language premorbidly, but only in early bilinguals; the effect of proficiency and use were instead limited for late bilinguals. In summary, although both AoA and proficiency appear to be relevant in shaping the bilingual brain, their relative role is not yet clear and the extent to which the proficiency level might scale down the role of AoA has not been yet investigated.

THE CURRENT META ANALYSIS

The principal aim of the present meta-analysis was to shed further light on the impact of AoA on the overall language brain representation and on those specifically associated with each language. Therefore, we tried to derive which brain regions bilinguals activate in a consistent way when performing language tasks in known languages. We carried out the analysis separately for the groups of early and late bilinguals. In addition, we wanted to inspect if some reliable activations could be found in a subgroup of very early bilinguals. After these more global analyses, we investigated the specific activations associated with each language, to then compare L1 and L2 within each group (i.e., early and late bilinguals), and L1s and L2s between groups. Lastly, we aimed to investigate the effect of proficiency. In particular, we inspected whether the two language networks in early and late bilinguals differed as the result of different proficiency levels.

With respect to the previous meta-analyses, this study: (i) explored more in depth the language networks associated with each language as the result of AoA first and of proficiency, second; (ii) investigated the language brain representation resulting from a very early L2 acquisition; and (iii) adopted quite stringent criteria for both paper inclusion and data analysis, in order to ascertain the strength of the resultant findings.

We hypothesized to confirm the results from previous meta-analyses in terms of an overall greater functional activation for the group of the late bilinguals with respect to that of the early bilinguals. Regarding the comparison between the two languages, we attended greater functional activation for L2 than L1 and expected to find this difference even in the group of early bilinguals. Finally, we expected the differences between early and late bilinguals and between L2 and L1 to reduce in high vs. low proficient bilinguals.

MATERIALS AND METHODS

Paper Search and Selection

In the current meta-analysis, we included the papers selected from the pool of English-written articles published between 1995 and the end of 2016. To be included, the papers had to report neuroimaging studies (by fMRI or PET) involving healthy adult participants (aged 18–60). We performed the research in MedLine and Scholar databases, using keywords such as "fMRI" or "functional MRI," "PET," "bilingual*." The sample was further integrated with some papers found by inspecting the list of references of the papers resulting from this research. The paper selection procedure is sketched out in the PRISMA flow chart (Moher et al., 2009) in **Figure 1**.

In the selection, we excluded cases of bimodal bilingualism (i.e., with one of the languages being a sign language) and studies assessing the language abilities specific to bilingualism, such as translation/interpretation and switching. We hence restricted the selection to the studies having addressed the main structural domains (i.e., lexico-semantics, phonology/articulation, and morpho-syntax) and we excluded those investigating more specific tasks, as the affective/emotional components of language (e.g., role of emotional words) or numbers and mathematics. The selection was not limited to specific language families. We were confident in including different language tasks from different languages in the same analysis, given that the algorithm we used (see afterwards) looks for the areas showing a convergence of activation across different experiments and therefore provides only consistently recurring activations. Finally, in order to reduce confounding effects, we also excluded studies performing assessments after learning/training processes (e.g., learning of new words or grammar rules, training in a barely mastered language) or after some manipulations to language exposure. Further, the participant samples in these studies were normally gender-balanced and quite homogeneous in terms of age (often including young adults). This assured more robust findings.

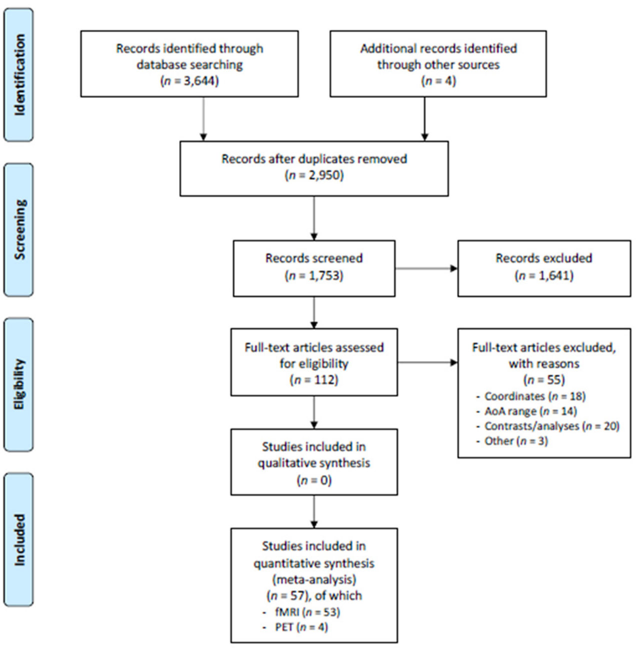


FIGURE 1 | PRISMA flowchart. Schematic representation of the paper search and selection process. From Moher et al. (2009).

For this analysis, we included only the studies that have been published after a peer-reviewed process. In this sense, the study might be subjected to a publication bias. Nevertheless, coordinate-based meta-analyses differ from effect-size meta-analyses in that they look for the spatial convergence between the reported coordinates; hence, they do not quantify the effect size, which is prone to bias. Therefore, this analysis seems to be less susceptible to region- and task-dependent biases and was not affected by the lacking inclusion of unpublished data (see Fox et al., 1998; Rottschy et al., 2012). Moreover, to reduce other sources of bias, we included only the results from whole-brain analysis, excluding those resulting from *a priori* selected ROIs.

This first selection resulted in 112 papers, which we further scrutinized to obtain the final sample. This selection was followed by the exclusion of some additional papers due to: (i) absent or incomplete (not full 3D) coordinates, including only coordinate ranges, coordinates that were reported only for single subjects, coordinates from *a priori* selected ROIs (not derived from the observed activations), $n = 19$; (ii) analyses where the contrasts were not informative (e.g., they did not differentiate between different languages or between bilinguals and monolinguals), were too specific or regarded a very low level of linguistic processing (e.g., passive viewing of single letters), *n* = 17; (iii) AoA that was not explicitly reported or did not fit our classification (see afterwards) *n* = 20; and (iv) other reasons (e.g., tasks assessing a linguistic ability "contaminated" by another aim, such as reading finalized to memorization), *n* = 3.

To define the two groups of early and late bilinguals, we adopted the age of 6 years as the AoA cutoff. The previously mentioned developmental steps occurring around this age guided our choice, further supported by the high number of studies having adopted this same cutoff. In fact, most of the studies classified bilinguals in early and late following an AoA that was, respectively, below and above 6 years of age. Alternatively, studies focused on either group, therefore either on early bilinguals, for instance from bilingual communities, or on late bilinguals, typically represented by people having learnt L2 at school. Concerning the very early bilinguals, an inspection of the studies we selected led us to include those with participants having an AoA up to 3 years. Unfortunately, the paucity of the studies on the early bilinguals that acquired L2 after L1 did not allow a specific analysis on this subgroup.

To meet the specific purposes of our paper, we, therefore, excluded the studies where AoA was not explicitly indicated or the reported AoA did not allow to include the participants in the groups we defined on the basis of the selected AoA cutoff (*n* = 17). Finally, in order to reduce additional sources of variability, we also excluded studies that investigated language learning in adulthood (*n* = 3).

Concerning proficiency, we observed that many studies reported self-rating assessments, or a general evaluation based on the performance in a single task (e.g., naming). Only a small percentage of studies reported a quantitative assessment by structured tests (e.g., TOEFL test for the English language). These ratings did not allow to reliably classify bilinguals from the proficiency viewpoint. Nevertheless, the studies in which the participants achieved a high score in a comprehensive language assessment or were defined to have a high proficiency, were, more consistently represented than those with low or intermediate proficiency. For this reason, we limited the analysis to the subsample of high proficient bilinguals and excluded from this subgroup the bilinguals whose proficiency in L2 was greater than in L1, in order to remove potential confounds.

The process of paper selection was preceded by the definition, by the three authors, of the objective criteria for study inclusion and exclusion. During the process, we consulted with one other to define additional criteria based on the issues that emerged in the meanwhile. At the end of the process, we discussed together about the residual papers that we did not know whether to include or not. In this way, we assured a consistent, unbiased selection procedure.

The final sample consisted of 57 papers (53 fMRI and four PET studies), from which we identified the groups of early bilinguals (74 experiments; 536 foci; 1,048 subjects), very early bilinguals (17 experiments; 91 foci; 227 subjects), and late bilinguals (174 experiments; 1,351 foci; 2,519 subjects), see **Supplementary Table S1** for paper list details.

Statistical Analyses

We carried out the meta-analyses using the coordinate-based activation likelihood estimation (ALE) algorithm developed for neuroimaging data (e.g., Turkeltaub et al., 2002; Eickhoff et al., 2009; Laird et al., 2009a,b). The algorithm looks for convergence across the experiment data, by evaluating whether the clustering is higher than that expected under the null distribution of a random spatial association. It, therefore, treats the reported foci as centers for 3D Gaussian probability distributions, to capture the spatial uncertainty associated with each focus. The provided probability distribution maps, which were weighted on the number of subjects in each study, described the probability for a given focus to lay within a given voxel.

We thresholded the probability maps for the main effect analyses at $p < 0.05$ (cluster-level corrected for multiple comparisons) and set a minimum cluster size to 200 mm^3 (25 voxels). For the analysis on individual languages, we reduced the extent threshold to 120 mm^3 (15 voxels). For the contrast analyses, we used threshold values of $p < 0.001$ (uncorrected) and a minimum cluster size of 80 mm^3 (10 voxels). Nevertheless, for the conjunction results, we retained only a minimum of 120-mm^3 (15 voxels) clusters, in order to exclude a possible incidental overlap between the ALE maps from individual analyses (see Rottschy et al., 2012).

We performed the following analyses:

(i) overall language brain representation in early, very early, and late bilinguals.
In the first preliminary analysis, we investigated the overall (not language-specific) functional brain representation of early and late bilinguals (main effects). We also performed the same analysis on a subgroup of early bilinguals that have acquired the two languages roughly simultaneously (up to the age of 3) and therefore defined as very early bilinguals.

(ii) L1 and L2 networks and between-language and between-group comparisons.
We then focused on the functional networks associated with each language. We performed the analysis separately for late and early bilinguals, excluding the very early bilinguals for whom a distinction between L1 and L2 based on the AoA was not possible. We first carried out the main effect analyses; next, we performed between-group analyses to compare the networks of L1s and L2s across the two groups and within-group analysis to compare the functional networks associated with L1 and L2 within each group.

(iii) L1 and L2 networks in proficient bilinguals and between-language and between-group comparisons.
We replicated the recently mentioned analyses on a subgroup of proficient bilinguals.

We reported the coordinates in the Montreal Neurological Institute (MNI) standard space. The coordinates that were standardized to the Talairach and Tournoux (1988) space in the included studies were converted to the MNI space by the icbm_spm2tal transform. To define the precise anatomical localization and label of the resulting areas, we used the SPM Anatomy toolbox (Eickhoff et al., 2005), running on MATLAB. We, therefore, reported the macro-anatomic localization and, when provided, the cytoarchitectonic location.

RESULTS

Whole Language Brain Representation in Early, Very Early, and Late Bilinguals

The main effect results for each group are reported in **Tables 1.1**, **1.2** and **Figure 2**.

Early Bilinguals

With regard to the early bilinguals, functional activations emerged in the following regions of the left hemisphere: (i) inferior parietal lobule (including the intraparietal sulcus—area hIP2); (ii) inferior occipital gyrus (i.e., fusiform area); (iii) precentral gyrus; (vi) rolandic operculum; and (v) inferior frontal gyrus (i.e., BA 44, BA 45, and the dorsolateral prefrontal cortex, DLPFC); right-sided activations included (vi) the cerebellum (lobule VIIa and crus I) and bilateral activations; (vii) middle temporal gyri (including the higher auditory cortex—area TE3); (viii) posterior-medial frontal gyri; and (ix) the insulae.

Very Early Bilinguals

The very early bilinguals displayed activation, in the left hemisphere, of the (i) middle temporal gyrus and, in both the hemispheres, of the (ii) cerebella (lobule VI, in the left hemisphere; lobule VIIa and crus I, in the right hemisphere).

Late Bilinguals

The late bilinguals' activation clusters included the following regions of the left hemisphere: (i) the inferior occipital gyrus (fusiform gyrus—area FG4); (ii) superior parietal lobule; (iii) middle temporal gyrus; (iv) precentral gyrus; (v) posterior-medial frontal gyrus; and (vi) inferior frontal gyrus (including BA 44, pars orbitaris, and DLPFC); activation clusters were also found in right (viii) angular gyrus (more precisely the intraparietal sulcus—area hIP3); and (ix) cerebellum (lobule VI, lobule VIIa, and crus I) and in bilateral (x) middle occipital gyrus (lateral cortex—area hOc4lp); and (xi) insulae.

L1 and L2 Networks in Early and Late Bilinguals

The functional brain activations associated with either L1 or L2 are detailed in **Tables 2**, **3** for L1 and L2, respectively and are all represented in **Figure 3**.

L1

Early bilinguals, when performing tasks in their L1, activated the following regions, all located in the left hemisphere: (i) the inferior temporal gyrus (fusiform gyrus); (ii) middle temporal

TABLE 1.1 | Main effect results of the activation likelihood estimation (ALE) meta-analysis for the groups of early and very early bilinguals.

Cluster (area)		MNI coordinates			Cluster size (voxels)	Extrema value
		x	*y*	*z*		
Early bilinguals						
1	L inferior parietal lobule	−24	−68	44	71	0.029
2	L inferior occipital (FG4)	−44	−58	−12	148	0.030
3	L middle temporal gyrus	−50	−48	4	41	0.025
5	L middle temporal gyrus (TE3)	−68	−32	2	152	0.037
6	R middle temporal gyrus (TE3)	70	−14	−8	53	0.029
7	L precentral gyrus	−46	−6	38	631	0.033
	L inferior frontal gyrus	−40	12	28		0.033
	L inferior frontal gyrus	−46	18	22		0.031
	L precentral gyrus	−42	0	28		0.027
8	L precentral gyrus	−52	2	50	25	0.026
9	L posterior-medial frontal gyrus	−2	2	66	264	0.036
	L posterior-medial frontal gyrus	−4	16	52		0.034
10	L rolandic operculum	−48	8	2	114	0.024
	L insula	−44	12	−4		0.022
	L inferior frontal gyrus (BA 44)	−56	8	10		0.021
11	R posterior-medial frontal gyrus	12	16	46	36	0.027
12	L insula	−30	18	4	124	0.028
	L insula	−32	26	0		0.027
13	R insula	36	24	−4	59	0.031
14	L inferior frontal gyrus (BA 45)	−54	30	4	137	0.032
15	R cerebellum (lobule VIIa, crus I)	36	−74	−28	25	0.023
Very early bilinguals						
1	L middle temporal gyrus	−66	−30	−2	46	0.020
2	L cerebellum (lobule VI)	−16	−68	−20	53	0.020
3	R cerebellum (lobule VIIa, crus I)	18	−68	−16	45	0.018

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 25 voxels. FG4, Fusiform area; hIP2, Horizontal tracts of the intraparietal sulcus; IPS, Intraparietal sulcus; TE3, Higher auditory cortex.

TABLE 1.2 | Main effect results of the ALE meta-analysis for the group of the late bilinguals.

Cluster (area)	MNI coordinates			Cluster size (voxels)	Extrema value
	x	*y*	*z*		
Late bilinguals					
1 L middle occipital gyrus (hOc4lp)	−28	−92	4	83	0.039
2 R middle occipital gyrus (hOc4lp)	36	−88	8	70	0.050
3 L inferior occipital gyrus (FG4)	−46	−64	−12	180	0.049
4 L superior parietal lobule	−26	−64	46	422	0.063
5 R angular gyrus (IPS, hIP3)	30	−62	48	63	0.039
6 L middle temporal gyrus	−52	−36	8	30	0.037
7 L inferior frontal gyrus	−44	12	28	2,630	0.091
L insula	−32	26	−2		0.071
L inferior frontal gyrus (BA 45)	−48	28	18		0.070
L inferior frontal gyrus (BA 44)	−54	12	10		0.062
L inferior frontal gyrus	−50	32	10		0.047
L inferior frontal gyrus (pars orbitaris)	−48	38	−8		0.039
L precentral gyrus	−50	10	50		0.039
L precentral gyrus	−52	2	50		0.033
L precentral gyrus	−48	−4	40		0.030
8 L posterior-medial frontal gyrus	−2	20	48	842	0.096
9 R insula	36	24	−4	159	0.058
10 R cerebellum (lobule VIIa, crus I)	36	−74	−26	25	0.035
11 R cerebellum (lobule VI)	22	−66	−22	43	0.037

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 25 voxels. FG4, Fusiform area; hIP2, hIP3, Horizontal tracts of the intraparietal sulcus; hOc4lp, Lateral occipital cortex; IPS, Intraparietal sulcus.

gyrus (area TE3); (iii) precentral gyrus; (iv) posterior-medial frontal gyrus; and (v) the inferior frontal gyrus (Broca's area—BA 45—and a region associable with the DLPFC).

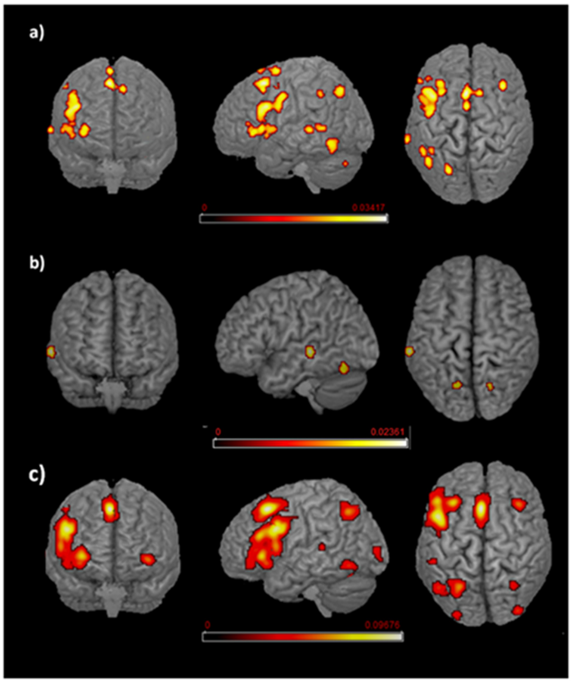


FIGURE 2 | Language networks associated with different age of appropriation (AoA). Rendered templates of the main effect analysis results for **(A)** early bilinguals, **(B)** very early bilinguals, and **(C)** late bilinguals. Color bars indicate the activation likelihood estimation (ALE) values.

The L1 activation clusters in late bilinguals included the following regions of the left hemisphere: (i) inferior occipital gyrus (fusiform gyrus—area FG4); (ii) middle temporal gyrus; (iii) precentral gyrus; (iv) posterior-medial frontal gyrus; (v) inferior frontal gyrus (including BA 44, BA 45, and DLPFC); and (vi) insula; right-sided activations were found in the (vii) superior-medial gyrus; and (viii) cerebellum (lobule VIIa and crus I).

L2

The activation clusters associated with early bilinguals' L2 emerged in the following regions, all in the left hemisphere: (i) the superior parietal lobule; (ii) precentral gyrus; (iii) inferior frontal gyrus (region including the DLPFC); and (iv) the posterior-medial frontal gyrus.

The late bilinguals' L2 functional activations were located in the following regions of the left hemisphere: (i) superior parietal lobule; (ii) inferior parietal lobule; (iii) superior temporal gyrus,; (iv) posterior-medial frontal gyrus; (v) inferior frontal gyrus (including BA 45, pars oribitaris, and DLPFC); and (vi) superior-medial gyrus; in the right hemisphere, activations emerged in (vii) calcarine gyrus (hOc1, V1); (viii) middle occipital gyrus (lateral cortex-area hOc4lp); (ix) angular gyrus; and (x) cerebellum (lobule VIIa and crus I); bilateral activations were observed in the (xi) insulae.

Between-Group Comparison Between L1s and L2s

The activation clusters resulting from the between-group contrast conditions (i.e., comparison between L1s and L2s across the two groups of early and late bilinguals) are reported in **Table 4** and **Figure 4**.

TABLE 2 | Results of the single ALE meta-analysis on L1 in the two groups of early and late bilinguals.

Cluster (area)		MNI coordinates			Cluster size (voxels)	Extrema value
		x	*y*	*z*		
L1: Early bilinguals						
1	L inferior temporal gyrus (FG4)	−46	−56	−12	58	0.027
2	L middle temporal gyrus (TE3)	−68	−34	2	107	0.034
3	L precentral gyrus	−50	0	32	15	0.020
4	L posterior-medial frontal gyrus	−2	4	64	35	0.024
5	L inferior frontal gyrus (BA 45)	−54	30	4	22	0.020
6	L inferior frontal gyrus	−46	32	10	58	0.020
L1: Late bilinguals						
1	L inferior occipital gyrus (FG4)	−46	−64	−12	400	0.028
2	L middle temporal gyrus	−58	−40	−2	184	0.027
	L precentral gyrus	−44	2	30		0.029
3	L precentral gyrus	−48	10	34	1,288	0.028
4	L inferior frontal gyrus (BA 44)	−54	10	8	424	0.029
5	L posterior-medial frontal gyrus	−4	20	48	1,504	0.047
6	L inferior frontal gyrus (BA 45)	−52	26	24	1,224	0.030
	L inferior frontal gyrus	−48	28	20		0.030
7	L insula	−28	28	−2	136	0.025
8	R superior medial gyrus	4	38	46	152	0.024
9	R cerebellum (lobule VIIa, crus I)	16	−90	−30	216	0.030

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 15 voxels. FG4, fusiform area; TE3, higher auditory cortex.

TABLE 3 | Results of the single ALE meta-analysis on L2 in the two groups of early and late bilinguals.

Cluster (area)		MNI coordinates			Cluster size (voxels)	Extrema value
		x	*y*	*z*		
L2: Early bilinguals						
1	L superior parietal lobule	−22	−70	46	22	0.023
2	L precentral gyrus	−42	−4	38	56	0.026
3	L inferior frontal gyrus	−48	18	22	56	0.027
4	L posterior-medial frontal gyrus	−4	20	66	18	0.022
L2: Late bilinguals						
1	R calcarine gyrus (V1)	14	−88	−2	32	0.027
2	R middle occipital gyrus (hOc4lp)	36	−88	8	29	0.028
3	R angular gyrus	28	−62	48	36	0.027
4	L superior parietal lobule	−24	−62	46	369	0.046
5	L inferior parietal lobule	−44	−40	42	31	0.029
6	L superior temporal gyrus	−54	−36	10	17	0.026
7	L inferior frontal gyrus	−46	12	26	2,212	0.078
	L insula	−32	26	−2		0.056
	L inferior frontal gyrus (BA 45)	−48	30	20		0.043
	L inferior frontal gyrus (BA 45)	−52	22	2		0.030
	L inferior frontal gyrus	−50	32	6		0.029
	L inferior frontal gyrus (pars orbitalis)	−52	24	−8		0.025
	L posterior-medial frontal gyrus	−2	22	50	815	0.062
8	L superior medial gyrus	−4	28	40		0.042
	L posterior-medial frontal gyrus	−4	8	54		0.037
9	R insula	38	24	−6	178	0.046
10	R cerebellum (lobule VIIa, crus I)	34	−72	−28	21	0.026

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 15 voxels. hOc4lp, lateral occipital cortex; V1, primary visual cortex.

L1: Conjunction (Early Bilinguals ∩ Late Bilinguals) and Subtraction Analyses (Early Bilinguals > Late Bilinguals and Late Bilinguals > Early Bilinguals)

No one area appeared to be consistently activated for L1 in conjunction of the two groups or in one group more than in the other.

L2: Conjunction (Early Bilinguals ∩ Late Bilinguals) and Subtraction Analyses (Early Bilinguals > Late Bilinguals and Late Bilinguals > Early Bilinguals)

Concerning L2, the areas activated in conjunction by the two groups were located in the left (i) inferior frontal gyrus (at the border between BA44 and DLPFC). The direct comparison

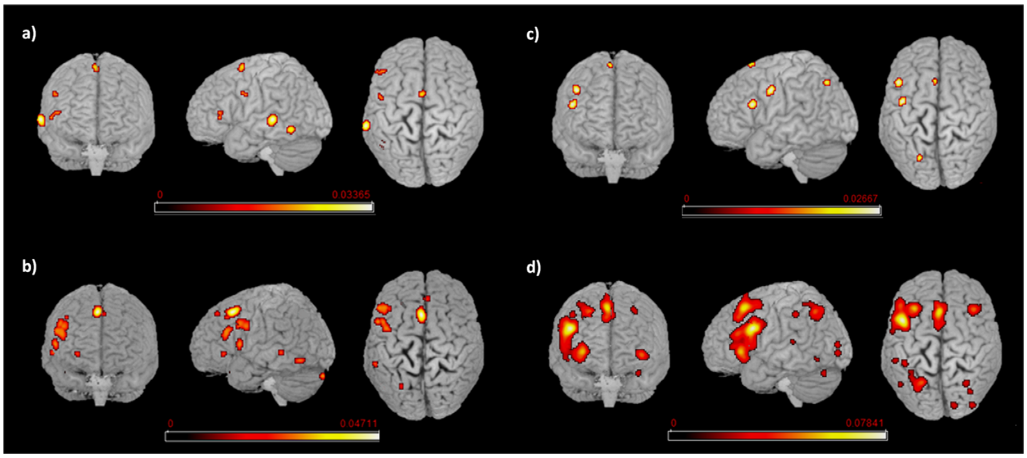


FIGURE 3 | Language networks associated with L1 and L2 in the two groups of early and late bilinguals. Rendered templates of the main effect results for **(A)** early bilinguals' L1; **(B)** late bilinguals' L1; **(C)** early bilinguals' L2; **(D)** late bilinguals' L2. Color bars indicate the ALE values.

TABLE 4 | Results of the between-group ALE meta-analysis for L1s and L2s.

Cluster (area)	MNI coordinates			Cluster size (voxels)	Extrema value
	x	*y*	*z*		
L1s					
Early bilinguals ∩ Late bilinguals					
No suprathreshold clusters of activation					
Early bilinguals > Late bilinguals					
No suprathreshold clusters of activation					
Late bilinguals > Early bilinguals					
No suprathreshold clusters of activation					
L2s					
Early bilinguals ∩ Late bilinguals					
1 L inferior frontal gyrus	−48	18	22	56	0.027
Early bilinguals > Late bilinguals					
No suprathreshold clusters of activation					
Late bilinguals > Early bilinguals					
No suprathreshold clusters of activation					

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.001$ (uncorrected) and 15 voxels for the conjunction analysis, and $p < 0.001$ (uncorrected) and 10 voxels for the subtraction analyses.

did not show any clusters activating more consistently in either group over the other.

Within-Group Comparison Between L1 and L2

Results of the within-group comparison are reported in **Table 5**.

Early Bilinguals: Conjunction (L1 ∩ L2) and Subtraction (L1 > L2 and L2 > L1) Analyses

For the early bilinguals' group, neither the conjunction nor the subtraction analysis provided suprathreshold activation clusters in the comparison between L1 and L2.

Late Bilinguals: Conjunction (L1 ∩ L2) and Subtraction (L1 > L2 and L2 > L1) Analyses

The late bilinguals activated the following left-hemisphere areas in conjunction with the two languages: (i) the precentral gyrus; (ii) posterior-medial frontal gyrus; and (iii) the inferior frontal gyrus (BA 45 and DLPFC).

The direct comparison between the two languages did not reveal any single region to be more consistently activated in L1 than in L2. Conversely, L2, when compared to L1, engaged more consistently in following regions, both in the left hemisphere: (i) the inferior frontal gyrus (region including the DLPFC); and (ii) the posterior-medial frontal gyrus.

L1 and L2 Networks in Proficient Bilinguals

We re-ran the previous analyses on a subgroup of highly proficient bilinguals (16 studies including the early bilinguals, 17 studies including the late bilinguals). The functional networks associated with either L1 or L2 are detailed in **Tables 6.1**, **6.2** for L1 and L2, respectively and are represented in **Figure 5**.

L1 in Proficient Bilinguals

For the early bilinguals, the functional activations associated with L1 emerged in the left:

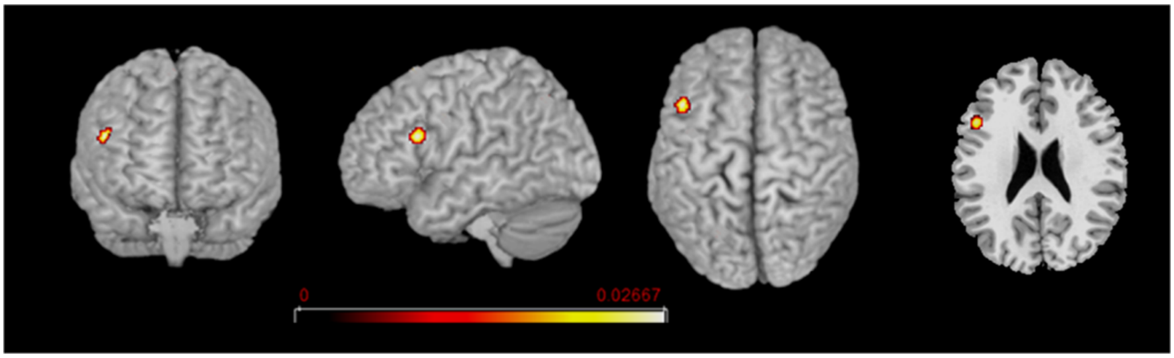


FIGURE 4 | L2 network comparison between the groups of early and late bilinguals. Rendered templates and axial projection of the conjunction analysis results for L2: Early bilinguals ∩ Late bilinguals. Color bars indicate the *Z*-score values.

TABLE 5 | Results of the within-group contrast ALE meta-analysis between L1 and L2 in the two groups of early and late bilinguals.

Cluster (area)	**MNI coordinates**			**Cluster size (voxels)**	**Extrema value**
	x	***y***	***z***		
Early bilinguals					
L1 ∩ L2					
No suprathreshold clusters of activation					
Early bilinguals, L1 > L2					
No suprathreshold clusters of activation					
Early bilinguals, L2 > L1					
No suprathreshold clusters of activation					
Late bilinguals					
L1 ∩ L2					
1 L precentral gyrus	−48	10	34	67	0.029
2 L precentral gyrus	−44	2	30		0.029
L posterior-medial frontal gyrus	−4	20	48	118	0.047
3 L inferior frontal gyrus (BA 45)	−52	26	24	74	0.030
L inferior frontal gyrus	−48	28	20		0.030
L1 > L2					
No suprathreshold clusters of activation					
L2 > L1					
1 L inferior frontal gyrus	−44	14	20	39	n/a
L inferior frontal gyrus	−43	12	24		n/a
2 L posterior-medial frontal gyrus	−2	20	58	17	n/a

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.001$ (uncorrected) and 15 voxels for the conjunction analysis, and $p < 0.001$ (uncorrected) and 10 voxels for the subtraction analyses.

TABLE 6.1 | Results of the single ALE meta-analysis on L1 in the two groups of proficient early and late bilinguals.

Cluster (area)	**MNI coordinates**			**Cluster size (voxels)**	**Extrema value**
	x	***y***	***z***		
L1: Proficient early bilinguals					
1 L middle temporal gyrus (TE3)	−68	−34	2	113	0.036
2 L posterior-medial frontal gyrus	−2	4	66	19	0.023
L1: Proficient late bilinguals					
1 L inferior frontal gyrus (BA 45)	−52	28	24	48	0.023

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 15 voxels. TE3, Higher auditory cortex.

(i) middle temporal gyrus (area TE3—higher auditory cortex); and (ii) the posterior-medial frontal gyrus.

In the late bilinguals, the functional network included the left: (i) posterior-medial frontal gyrus; and (ii) the inferior frontal gyrus (BA 45).

TABLE 6.2 | Results of the single ALE meta-analysis on L2 in the two groups of proficient early and late bilinguals.

Cluster (area)		MNI coordinates			Cluster size (voxels)	Extrema value
		x	*y*	*z*		
L2: Proficient early bilinguals						
1	L inferior frontal gyrus	−48	18	22	56	0.027
L2: Proficient late bilinguals						
1	L calcarine gyrus (hOc1)	−8	−96	0	26	0.023
2	L inferior parietal lobe	−28	−70	48	120	0.028
3	L middle temporal gyrus (TE3)	−66	−24	−2	31	0.023
4	L caudate nucleus	−8	8	0	16	0.022
5	L inferior frontal gyrus	−46	12	26	549	0.040
	L inferior frontal gyrus (BA 45)	−50	26	22		0.026
6	L posterior-medial frontal gyrus	−2	22	50	519	0.047
	L posterior-medial frontal gyrus	−4	20	66		0.023
7	L inferior frontal gyrus	−52	24	−8	37	0.023
8	L insula	−32	26	−2	230	0.038
9	R caudate nucleus	12	18	0	15	0.021
10	R insula	38	26	−4	102	0.035

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.05$ (corrected) and extent threshold of 15 voxels. hOc1, Primary visual cortex (V1); TE3, Higher auditory cortex.

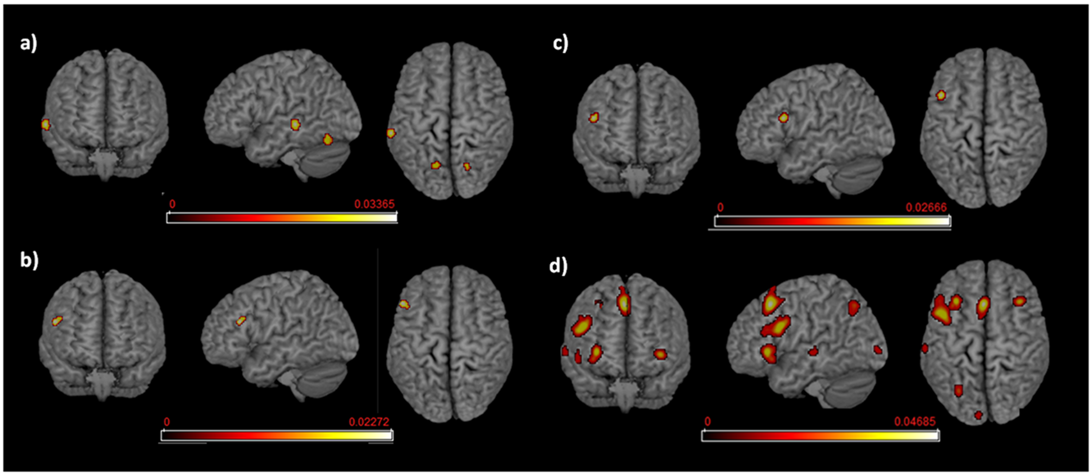


FIGURE 5 | Language networks associated with L1 and L2 in the groups of proficient early and late bilinguals. Rendered templates of the main effect results for **(A)** proficient early bilinguals' L1; **(B)** proficient late bilinguals' L1; **(C)** proficient early bilinguals' L2; **(D)** proficient late bilinguals' L2. Color bars indicate the ALE values.

L2 in Proficient Bilinguals

The early bilinguals' L2 significantly activated a portion of the left (i) inferior frontal gyrus (region including the DLPFC).

The late bilinguals' functional activations associated with L2 included different areas in the left hemisphere: (i) inferior parietal cortex; (ii) inferior frontal gyrus (including BA 45 and a region associable with the DLPFC); and (iii) posterior-medial frontal gyrus; bilateral activation was found in the (iv) caudate nuclei and (v) insulae.

Between-Group Comparison Between Early and Late Proficient Bilinguals

Results of the between-group comparison are reported in **Table 6.3**.

L1: Conjunction (Early Bilinguals ∩ Late Bilinguals) and Subtraction (Early Bilinguals > Late Bilinguals and Late Bilinguals > Early Bilinguals) Analyses

For L1, neither the conjunction nor the subtraction analysis provided suprathreshold activation clusters in the comparison between L1 and L2.

L2: Conjunction (Early Bilinguals ∩ Late Bilinguals) and Subtraction (Early Bilinguals > Late Bilinguals and Late Bilinguals > Early Bilinguals) Analyses

For L2, the conjunction analysis provided a shared activation cluster between early and late bilinguals in the left (i) inferior frontal gyrus (region including the DLPFC).

In the subtraction analyses, suprathreshold activation clusters did not result from either comparison.

TABLE 6.3 | Results of the between-group ALE meta-analysis for L1s and L2s in proficient bilinguals.

Cluster (area)	MNI coordinates			Cluster size (voxels)	Extrema value
	x	*y*	*z*		
Proficient bilinguals' L1s					
Early bilinguals ∩ Late bilinguals					
No suprathreshold clusters of activation					
Early bilinguals > Late bilinguals					
No suprathreshold clusters of activation					
Late bilinguals > Early bilinguals					
No suprathreshold clusters of activation					
Proficient bilinguals' L2s					
Early bilinguals ∩ Late bilinguals					
L inferior frontal gyrus	−48	18	22	56	0.027
Early bilinguals > Late bilinguals					
No suprathreshold clusters of activation					
Late bilinguals > Early bilinguals					
No suprathreshold clusters of activation					

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.001$ (uncorrected) and 15 voxels for the conjunction analysis, and $p < 0.001$ (uncorrected) and 10 voxels for the subtraction analyses.

TABLE 6.4 | Results of the within-group ALE meta-analysis for L1s and L2s in proficient bilinguals.

Cluster (area)	MNI coordinates			Cluster size (voxels)	Extrema value
	x	*y*	*z*		
Proficient early bilinguals					
L1 ∩ L2					
No suprathreshold clusters of activation					
Early bilinguals, L1 > L2					
No suprathreshold clusters of activation					
Early bilinguals, L2 > L1					
No suprathreshold clusters of activation					
Proficient late bilinguals					
L1 ∩ L2					
1 L inferior frontal gyrus (BA 45)	−54	28	20	32	0.023
L1 > L2					
No suprathreshold clusters of activation					
L2 > L1					
No suprathreshold clusters of activation					

Note. For the anatomical localization the macro-anatomic area is indicated and, when provided, the cytoarchitectonic location is indicated (in parentheses). Analyses were performed with a threshold of $p < 0.001$ (uncorrected) and 15 voxels for the conjunction analysis, and $p < 0.001$ (uncorrected) and 10 voxels for the subtraction analyses.

Within-Group Comparison Between L1 and L2 in Proficient Bilinguals

Results of the within-group comparison are reported in **Table 6.4**.

Early Proficient Bilinguals: Conjunction (L1 ∩ L2) and Subtraction (L1 > L2 and L2 > L1) Analyses

For the early bilinguals' group, neither the conjunction nor the subtraction analysis provided suprathreshold activation clusters in the comparison between L1 and L2.

Late Proficient Bilinguals: Conjunction (L1 ∩ L2) and Subtraction (L1 > L2 and L2 > L1) Analyses

For the late bilinguals' group, the conjunction analysis showed a shared activation cluster between L1 and L2 in the left (i) inferior frontal gyrus (BA 45).

In the subtraction analyses, neither comparison provided suprathreshold activation clusters.

DISCUSSION

The present meta-analysis aimed to inspect whether AoA and the traditional classification in early and late bilinguals have an actual role in shaping the bilingual language brain networks, even when accounting for the level of proficiency. We hence identified the two groups of early and late bilinguals (by taking 6 years of age as the AoA cutoff), and also a subgroup of very early bilinguals, in order to investigate the effect of the simultaneous acquisition of two languages. The first preliminary analyses were comprehensive of both the languages the participants knew, as we wanted to obtain a global overview of the whole language network in the three groups. We then performed more focused analyses to assess the functional networks specifically associated with each language, and between- and within-group comparisons between the languages. In this way, we inspected whether, irrespective of other intervening factors, the conventional classification in early and late bilinguals reflected actual differences in the related brain networks. Finally, we replicated these analyses by including only the highly proficient bilinguals, in order to check whether AoA was still relevant when proficiency was comparable (and high) between early and late bilinguals. We carried out these language-specific analyses only for the late bilinguals and for the early bilinguals for which the identification of the first and second language was possible.

Early, Very Early, and Late Bilinguals

As a first account, we showed that both early and late bilinguals displayed a widespread language network, which was located predominantly in the left (dominant) hemisphere. This network included the classical language areas, together with additional cortical and subcortical regions possibly recruited to support the language functions. For instance, in line with the monolingual language network, both early and late bilinguals activated the classical language areas, such as the Broca's area (BA 44 and BA 45) known to be involved in a variety of language domains (for reviews see Grodzinsky and Santi, 2008; Friederici, 2011; Price, 2012). Additional shared activations emerged in the left premotor cortex (precentral gyrus) and pre-SMA (posterior-medial frontal gyrus). These regions are involved in articulation-related processes (e.g., Hickok and Poeppel, 2004; Indefrey and Levelt, 2004; Alario et al., 2006; Kemeny et al., 2006), but also in other language tasks, including phonological rehearsal (e.g., Démonet et al., 1992; Paulesu et al., 1993; Awh et al., 1996). However, the role of pre-SMA seems to go beyond these functions to include the control in language use. Actually, Abutalebi and Green (2007, 2016) proposed this area to be one of the stations of the language control network (see afterwards).

Other activation clusters included the middle temporal gyrus—particularly the area associated with the auditory cortex—and fusiform gyrus, both in the left hemisphere. The former is known to be specialized in the perception of words over other non-linguistic sounds (e.g., Binder et al., 1997). The fusiform gyrus is specific for the recognition of words as well, in particular in their written form, across diverse languages and scripts (e.g., Cohen et al., 2000; Turkeltaub et al., 2002; Vigneau et al., 2005; Price and Devlin, 2011). The activation of these two regions might reflect the nature of the stimuli, either auditory or written. Nevertheless, the fusiform gyrus was also shown to contribute to lexical-semantic access, by working in association with the other areas of the middle and inferior temporal gyri (e.g., Papathanassiou et al., 2000; Démonet et al., 2005).

Besides the classical fronto-temporal language areas, also the parietal lobe—particularly the posterior parietal cortex (PPC)—activated in both bilingual groups. This region is not typically devoted to language, although some studies reported its involvement in the performance of some language tasks (for instance in vocabulary learning, see Pasqualotto et al., 2015). Interestingly, this region is relevant to working memory and its activation might, therefore, reflect the heightened necessity for the bilingual speakers to reinforce and elaborate the linguistic information associated with each language (e.g., Gold et al., 2005; Hartwigsen et al., 2010).

Early and late bilinguals also activated the right cerebellum, which is reciprocally connected with the left neocortex and whose involvement in language is becoming progressively more apparent (see for an overview De Smet et al., 2013; Mariën et al., 2014).

Both bilingual groups activated brain areas that more likely reflect the act of having to handle more than one language. In particular, we observed a prominent activation in the DLPFC, which is traditionally associated with high cognitive (executive) functions (e.g., Daffner et al., 2000; McDonald et al., 2000). With respect to bilinguals, DLPFC has been proposed to be a chief station of the network that regulates language selection and control; this region hence modulates the use of each language, for instance by inhibiting the interfering one (i.e., not-in-use; e.g., Abutalebi and Green, 2007, 2016). Another key area of this network has been proposed to be the pre-SMA, which was activated as well, as previously discussed.

An analogous functional interpretation can also be proposed for the insula, which was activated in both hemispheres. Although this region is traditionally viewed as part of the limbic system, its role in language is becoming progressively more evident. Regarding bilingualism, previous evidence supported its involvement in the mechanisms of switching and control (e.g., Wager et al., 2005). Despite the greater role attributed to the left insula, several studies reported bilateral activations in relation to diverse language functions, both receptive and expressive (see the meta-analysis by Oh et al., 2014). Nevertheless, the specific involvement of the right insula requires further investigation. However, it is generally thought to support the dominant hemisphere in various language functions, especially when they become cognitively more demanding (see also the meta-analysis by Vigneau et al., 2011), as can occur in bilingual settings.

Although we did not carry out a direct comparison with the monolingual network (as only a small percentage of studies included data on monolinguals), these preliminary results supported the notion that "the bilingual is not two monolinguals in one person" (Grosjean, 1989); this means that the language network in bilinguals is different from the one that could result from the sum of two language-specific networks in monolinguals (Fabbro, 1999). Actually, bilinguals have to constantly regulate the use of a certain language even when immersed in a monolingual mode, as both languages are, to a certain extent, active (e.g., Marian and Spivey, 2003; Dijkstra, 2005). Even when only one language is in use, there is a continuous interference from the other language, which therefore has to be inhibited (see Paradis, 2004).

The further analysis we conducted investigated the effects associated with an almost concurrent acquisition of the two languages. The functional activations in the very early bilinguals' group were found in a few areas, such as the left middle temporal gyrus and bilateral cerebella. Interesting was the activation of the left cerebellum, which did not emerge from the previous analyses. This finding leads us to stress once more of the importance of this subcortical structure and hints at a speculative hypotheses for its role (see Ullman, 2006; Paradis, 2009; and **Supplementary Material** for details).

The fact that very early bilinguals activated in a consistent manner in only a few regions could reflect two possible reasons. First, it is reasonable that these bilinguals need the recruitment of a lower number of regions to perform the language tasks because the very precocious acquisition could imply a lower cognitive effort. This is only a partial explanation, given that the resultant activation clusters did not include other relevant areas of the language network. Hence, this finding may also reflect the low number of studies that have addressed very early language acquisition and, consequently, the low number of provided foci (see the "Materials and Methods" section). These analyses, therefore, need to be replicated once a suitable number of studies is available.

L1 and L2 Brain Representation

Whereas the previous analyses provided a general overview of the overall brain functioning in response to different AoA, the subsequent analyses were devoted to the investigation of the language brain activations associated with each language. Because a distinction between L1 and L2 in very early bilinguals was rarely possible, we carried out this investigation separately for late bilinguals and for the early bilinguals for which such distinction was achievable.

L1

Concerning L1, the results showed different functional networks for early and late bilinguals. With regards to the former, activations (all left-sided) emerged in the classical language areas (i.e., fusiform gyrus, middle temporal gyrus, precentral gyrus, and BA 45) and in regions devoted to cognitive control (i.e., pre-SMA and DLPFC). This suggests that, even in early bilinguals and even when dealing with the first language, there is the need to control and regulate the language use, by possibly suppressing the activation of the second language, which is likely to exert a strong interference.

Regarding late bilinguals, a greater number of activated clusters emerged. These included language-associated areas (as the fusiform gyrus, the middle temporal gyrus, the precentral gyrus, and Broca's area), control areas (i.e., DLPFC, pre-SMA, and insula in the left hemisphere, and ACC in the right), and the right cerebellum. Also, in late bilinguals, then, language control seems to be required even when performing tasks in L1. This could occur because the first language has to be strongly inhibited in a bilingual context because it tends to prevail even when it is the L2 being used. Consequently, when L1 has to be activated again, great cognitive resources are required to overcome this inhibition, thus implying increased cognitive effort (see switching studies, e.g., Meuter and Allport, 1999; Garbin et al., 2011).

L2

Also regarding L2, the network of activations was more substantial for the group of late bilinguals compared to early bilinguals (who activated the superior parietal lobule, precentral gyrus, DLPFC, and pre-SMA), in part probably because of the lower number of contrasts associated with the latter. The late bilinguals' functional activations were widespread and spanned from the left parietal lobe—both inferior and superior—to the left superior temporal gyrus, frontal regions specifically devoted to language (i.e., BA 45 and pars orbitalis) or control (i.e., left pre-SMA, ACC, and DLPFC, and bilateral insuale), and to the right cerebellum. Some clusters of activation also emerged in the right hemisphere and concerned posterior areas located in the occipital cortex and angular gyrus.

Findings on the number and extent of activations observed for late-learned L2 were not surprising and support the hypothesis of a greater involvement of the areas typically associated with language (e.g., wider activations to compensate for lower efficiency), those devoted to control, and the additional involvement of the right hemisphere. Concerning the activations in the left inferior parietal lobule and in the right hemisphere, a detailed inspection of the contrasts, having concurred to these clusters, helped us hypothesize the rationale for their involvement. In the **Supplementary Material**, we illustrated these hypotheses and stressed the role that some brain areas, especially in the parietal lobe, can hold in some language functions under specific conditions.

Between- and Within-Group Comparison Between L1 and L2

The last analyses we performed were aimed at comparing L1 and L2 within and between each group. For the between-group analysis, L1 did not appear to determine any specific activation for either early or late bilinguals. Similarly, the subtraction analysis did not provide group-specific activations either. However, the main effect results revealed a higher number of activation clusters for the late bilinguals' group; we cannot exclude this finding to depend on the higher number of available contrasts and the consequent higher probability to draw consistent activations. This limitation prevents us from stating with certainty whether—and in case with which extent—the L1 network might differ as the result of a different AoA. In other words, we cannot comment on the possible feedback impact of this factor on the L1 brain representation (e.g., Titone et al., 2011, for eye-tracking findings).

With regards to the comparison between early and late bilinguals' L2s, the sole area that activated in conjunction with the two groups, possibly indicating its role irrespective of AoA, was a portion of the left IFG at the border between Broca's area and DLPFC. The lack of additional common clusters can reasonably reflect the paucity of consistent activations in the group of early bilinguals. In addition, this could also result from actual differences in the L2 activation sites, which can be located in close areas but peak at different coordinates (see Indefrey, 2006). With respect to the direct subtractions between the two groups, the lack of any significant specific activation was not surprising, again possibly resulting from high data variability and consequent lack of consistency. Liu and Cao (2016), who adopted more lenient threshold parameters, reported the findings from a similar subtraction analysis in which they compared the specific L2 activations between early and late bilinguals. Comparable to our analyses, they did not find any region that was more consistently activated in the early bilinguals' group. They did, however, find a specific late bilinguals' activation in the region of the left superior frontal gyrus, showing again the greater recruitment of executive control regions as the result of late L2 learning.

Moreover, we can tentatively attribute the lack of consistent results to the inadequacy of the sharp AoA-based classification in reflecting the bilingual brain development. In fact, although development is characterized by clearly defined steps, changes in language appropriation flexibility are not expected to reduce sharply, but rather gradually. In this sense, the re-conceptualization of the critical period in the *sensitive* period could better account for the gradualness of the process (e.g., Flege et al., 1999). In relation to this, it is also important to underline that the critical age for language appropriation was also shown to depend on the language domain. For instance, an early appropriation seems to be more crucial for grammar and phonology/articulation, whereas late appropriation has a less negative impact on lexico-semantics (Paradis, 1994, 2009; Ruben, 1997). In this study, we purposefully investigated the role of AoA on the whole L1 and L2 brain representation, therefore independently from the language domain. A previous meta-analysis, however, inspected the language networks associated with lexico-semantics (Indefrey, 2006), which is probably the most studied domain. This analysis was exploratory as it included a small sample of studies. With the increasing number of neuroimaging studies in bilinguals, in future years it will be possible to have a suitable number of studies in each language domain to investigate the related networks in the two languages.

With regard to the within-group comparisons in the group of the early bilinguals, neither the conjunction nor the subtraction analyses between L1 and L2 provided significant findings. The stringent criteria of paper selection together with the relatively conservative thresholding parameters could possibly explain the discrepancy with Liu and Cao (2016) findings, which showed more relevant L2 activations in the left frontal cortex and insula.

Concerning the late bilinguals, our conjunction analysis revealed L1 and L2 to activate common sites in both classical language areas (i.e., left precentral gyrus and BA 45) and in regions supporting general executive functions (i.e., left pre-SMA and DLPFC). The comparison between the two languages did not provide any regions that appeared to activate selectively for L1. On the contrary, and in line with previously discussed findings, L2 appeared to activate the left pre-SMA and DLPFC more robustly, in a close location to that which emerged from both the conjunction analysis and the same subtraction analysis reported in Liu and Cao (2016).

Trying to interpret these results in the light of the clinical findings on the bilingual patients with aphasia is quite tricky. Clinical literature indeed reports a plethora of different cases, in which the two languages were comparably affected (parallel aphasia) or not (differential aphasia); further, in the latter case, the most affected language could be represented by either L1 or L2.

In this respect, some clinical findings support the role of AoA, by reporting higher impairment in the language that had been learned late (e.g., Diéguez-Vide et al., 2012). Nevertheless, the variety of clinical profiles indicates that many are the factors that contribute to the language brain representation and possible impairment. Among these factors, proficiency and use/exposure have a relevant role in determining which language could be more affected by a clinical event (e.g., Gray and Kiran, 2013). This means that the language that was highly mastered prior to the brain injury is likely to be more resistant to damage, and could, therefore, be better preserved (e.g., Samar and Akbari, 2012). However, AoA has been proposed to retain a leading role, with the role of performance instead emerging only when both languages have been learnt early (see Kuzmina et al., 2019). In the current study, we, therefore, inspected whether, after having accounted for proficiency, AoA could still account for differences between the two languages. In other words, we wanted to assess the actual role of proficiency, which also emerged from a meta-analysis on healthy individuals, in which, however, the role of AoA was not accounted for (Sebastian et al., 2011).

Proficient Bilinguals

The last analyses we performed aimed to investigate the language brain representation in early and late bilinguals by removing possible confounding effects due to proficiency. For methodological reasons, we could perform the analyses only on the proficient bilinguals; as long as a high proficiency level was expected to reduce the cognitive effort associated with L2, we inspected whether, proficiency held constant, different brain activations still emerged as the result of different AoA.

These analyses as well were almost exploratory. In fact, the number of experiments included in each analysis was rather low (except for the late bilinguals' L2) and this was probably the reason why the "classical" language network could not be traced and only a few activation clusters resulted even from the main effect analyses. Nevertheless, this factor, together with the application of stringent thresholds, probably provided the most robust activation clusters for the inspected conditions, which are therefore expected to be highly reliable.

Concerning L1, the main effect analysis showed, in both groups, left-sided activation clusters in areas typically involved in language (in the middle temporal gyrus in early bilinguals and in BA 45 in late bilinguals). Further, the early bilinguals also activated the left pre-SMA. Although our data only allow for speculative interpretations, these results seem to suggest that handling L1 as well requires a certain cognitive control, even when proficiency in both languages is high and appropriation occurred at an early age. This indicates the constant need for bilinguals to monitor and regulate the use of both languages (Abutalebi and Green, 2007, 2016; Grosjean and Li, 2013).

With respect to L2, it is interesting to note that the language brain representation in the late bilinguals was consistently wider than that of the early bilinguals and that the two groups shared a cluster in the left inferior frontal gyrus, at the border between BA 45 and DLPFC. Nevertheless, no one cluster resulted from the direct comparison between the two groups. However, the main effect analysis in the late bilinguals' group provided activation clusters that did not result from the previous analysis and that corresponded to the bilateral caudate nuclei, one of the regions included in the Abutalebi and Green's (2007, 2016) language control network. An overall observation, therefore, suggests that, in spite of AoA and proficiency, L2 nevertheless requires the involvement of the executive functions, although the cognitive load appeared to be much greater when appropriation occurred after the age of 6. Late bilinguals, for instance, activated the insula in both hemispheres and control areas such as the pre-SMA. However, we have to remember that these findings might reflect, at least in part, the lower number of contrasts included in the early bilinguals' analysis.

CONCLUSION

Overall, results from these several meta-analyses lead us to conclude that, globally, bilinguals performing language tasks habitually recruit some additional brain regions with respect to the classical language network areas. These additional regions are involved in general cognitive functions, suggesting the constant effort experienced by every bilingual to manage the two languages. Even L1 and every other language possibly acquired since early childhood seem to call for this control. When dealing with two languages, there is clearly a need for their coordination, with the constant inhibition of the not-in-use language (Fabbro, 1999; Abutalebi and Green, 2007; Grosjean and Li, 2013). Nevertheless, in agreement with previous findings, we generally observed that the cognitive effort is stronger for L2, especially when this was learned late (e.g., Indefrey, 2006). The cognitive effort appeared to be present even in proficient bilinguals, although proficiency is likely to reduce the cognitive load associated with late L2 appropriation. This indicates that an early vs. late AoA significantly shapes the bilingual brain, although high proficiency can modulate the languages' functional representation (Fabbro and Cargnelutti, 2018).

The involvement of general cognitive areas is also, from a clinical viewpoint, a relevant finding. Actually, the cases of differential bilingual aphasia (where one language is more affected than the other) have also been explained in terms of control difficulties (e.g., Verreyt et al., 2013) and rehabilitation programs also focusing on the general cognitive functions were observed to promote language recovery after a brain insult (e.g., Hillis, 2001).

In this study, we did not carry out analyses for the main language domains separately, first because our aim was to identify the most relevant brain regions independently from the assessed task, and, second, because there was not an adequate number of studies to be included in these separate analyses. However, as the different domains are expected to rely more on either AoA (i.e., morpho-syntax and phonology/articulation) or proficiency (i.e., lexico-semantics), future analyses should investigate how these factors modulate the brain representation of these domains.

Further, for reasons we have explained, we found lower-than-expected significant activations in the comparison between the two languages. Intraoperative stimulation mapping studies in bilingual patients showed that the two languages shared many language sites, whereas other sites appeared to be language-specific (e.g., Roux and Trémoulet, 2002; Roux et al., 2004). However, there was a certain inter-subject variability, which could not be attributed uniquely to the different patients' language history (e.g., AoA or proficiency). These reports result from a clinical condition that might have induced brain reorganization processes and we cannot, therefore, make a direct comparison with our findings. Nevertheless, they suggest the complex interplay between the diverse factors in shaping the language brain representation in bilingual people. Factors such as the linguistic distance between the two known languages, the educational level, and possibly gender could modulate the differential representation of L1 and L2. Future studies should also address proficiency and other relevant parameters, including language exposure and use, in a more thorough way, in order to allow for a reliable assessment of their role. This, in turn, will contribute to a better understanding of the clinical reports of parallel and differential impairment and would, therefore, contribute to the rehabilitation program setting.

AUTHOR CONTRIBUTIONS

All the authors contributed to study design, data analysis, result interpretation, and manuscript drafting.

FUNDING

The project has been funded by the Province of Udine (Italy). The Cognitive Neuroscience Laboratory, DILL, University of Udine (Università degli Studi di Udine) will support the publication fee.

ACKNOWLEDGMENTS

We would like to thank the Province of Udine (Department of Culture), which financially supported the realization of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00154/full#supplementary-material

REFERENCES

Abutalebi, J., and Green, D. W. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003

Abutalebi, J., and Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Biling. Lang. Cogn.* 19, 689–698. doi: 10.1017/s1366728916000225

Alario, F. X., Chainay, H., Lehericy, S., and Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Res.* 1076, 129–143. doi: 10.1016/j.brainres.2005.11.104

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., and Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: evidence from PET. *Psychol. Sci.* 7, 25–31. doi: 10.1111/j.1467-9280.1996.tb00662.x

Berken, J. A., Gracco, V. L., Chen, J.-K., Watkins, K. E., Baum, S. M., Callahan, M., et al. (2015). Neural activation in speech production and reading aloud in native and non-native languages. *Neuroimage* 112, 208–217. doi: 10.1016/j.neuroimage.2015.03.016

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362. doi: 10.1523/jneurosci.17-01-00353.1997

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., and Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain Lang.* 120, 282–289. doi: 10.1016/j.bandl.2011.09.003

Callan, D. E., Callan, A. M., and Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Front. Neurosci.* 8:275. doi: 10.3389/fnins.2014.00275

Callan, D. E., Jones, J. A., Callan, A. M., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22, 1182–1194. doi: 10.1016/j.neuroimage.2004.03.006

Cao, F., Kim, S. Y., Liu, Y., and Liu, L. (2014). Similarities and differences in brain activation and functional connectivity in first and second language reading: evidence from Chinese learners of english. *Neuropsychologia* 63, 275–284. doi: 10.1016/j.neuropsychologia.2014.09.001

Cao, F., Tao, R., Liu, L., Perfetti, C. A., and Booth, J. R. (2013). High proficiency in a second language is characterized by greater involvement of the first language network: evidence from Chinese learners of English. *J. Cogn. Neurosci.* 25, 1649–1663. doi: 10.1162/jocn_a_00414

Casey, B. J., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol. Psychol.* 54, 241–257. doi: 10.1016/s0301-0511(00)00058-2

Chan, A. H., Luke, K.-K., Li, P., Yip, V., Li, G., Weekes, B., et al. (2008). Neural correlates of nouns and verbs in early bilinguals. *Ann. N Y Acad. Sci.* 1145, 30–40. doi: 10.1196/annals.1416.000

Chee, M. W., Hon, N., Lee, H. L., and Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *Neuroimage* 13, 1155–1163. doi: 10.1006/nimg.2001.0781

Chee, M. V. L., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J., et al. (2000). Overlap and dissociation of semantic processing of Chinese characters, English words and pictures: evidence from fMRI. *Neuroimage* 12, 392–403. doi: 10.1006/nimg.2000.0631

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Henaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307. doi: 10.1093/brain/123.2.291

Cook, M., Murdoch, B. E., Cahill, L., and Whelan, B.-M. (2004). Higher-level language deficits resulting from left primary cerebellar lesions. *Aphasiology* 18, 771–784. doi: 10.1080/02687030444000291

Daffner, K. R., Mesulam, M. M., Scinto, L. F., Acar, D., Calvo, V., Faust, R., et al. (2000). The central role of PFC in directing attention to novel events. *Brain* 123, 927–939. doi: 10.1093/brain/123.5.927

Das, T., Padakannaya, P., Pugh, K. R., and Singh, N. C. (2011). Neuroimaging reveals dual routes to reading in simultaneous proficient readers of two orthographies. *Neuroimage* 54, 1476–1487. doi: 10.1016/j.neuroimage.2010.09.022

De Bleser, R., and Kauschke, C. (2003). Acquisition and loss of nouns and verbs: parallel or divergent patterns? *J. Neurolinguist.* 16, 213–229. doi: 10.1016/S0911-6044(02)00015-5

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain* 115, 1753–1768. doi: 10.1093/brain/115.6.1753

Démonet, J. F., Thierry, G., and Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. *Physiol. Rev.* 85, 49–95. doi: 10.1152/physrev.00049.2003

De Smet, H. J., Paquier, P., Verhoeven, J., and Mariën, P. (2013). The cerebellum: its role in language and related cognitive and affective functions. *Brain Lang.* 127, 334–342. doi: 10.1016/j.bandl.2012.11.001

Diéguez-Vide, F., Gich-Fullà, J., Puig-Alcántara, J., Sánchez-Benavides, G., and Peña-Casanova, J. (2012). Chinese-spanish-catalan trilingual aphasia: a case study. *J. Neurolinguist.* 25, 630–641. doi: 10.1016/j.jneuroling.2012.01.002

Dijkstra, T. (2005). "Bilingual word recognition and lexical access," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. De Groot (Oxford: Oxford University Press), 179–201.

Ding, G., Perry, C., Peng, D., Ma, L., Li, D., Xu, S., et al. (2003). Neural mechanisms underlying semantic and orthographic processing in Chinese-English bilinguals. *Neuroreport* 14, 1557–1562. doi: 10.1097/01.wnr.0000087906.78892.8e

Druks, J., and Weekes, B. S. (2013). Parallel deterioration to language processing in a bilingual speaker. *Cogn. Neuropsychol.* 30, 578–596. doi: 10.1080/02643294.2014.882814

Edmonds, L., and Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *J. Speech Lang. Hear. Res.* 49, 729–748. doi: 10.1044/1092-4388(2006/053

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

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Fabbro, F. (1999). *The Neurolinguistics of Bilingualism.* Hove: Psychology Press.

Fabbro, F. (2001). The bilingual brain: cerebral representation of languages. *Brain Lang.* 79, 211–222. doi: 10.1006/brln.2001.2481

Fabbro, F., and Cargnelutti, E. (2018). *Neuroscienze del Bilinguismo (Neuroscience of Bilingualism).* Roma: Astrolabio.

Feng, G., Chen, H.-C., Zhu, Z., He, Y., and Wang, S. (2015). Dynamic brain architectures in local brain activity and functional network efficiency associate with efficient reading in bilinguals. *Neuroimage* 119, 103–118. doi: 10.1016/j.neuroimage.2015.05.100

Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (1999). Age constraints on second-language acquisition. *J. Mem. Lang.* 41, 78–104. doi: 10.1006/jmla.1999.2638

Fox, P. T., Parsons, L. M., and Lancaster, J. L. (1998). Beyond the single study: function/location metanalysis in cognitive neuroimaging. *Curr. Opin. Neurobiol.* 8, 178–187. doi: 10.1016/s0959-4388(98)80138-4

Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J., and Viallet, F. (2005). Articulation in early and late bilinguals' two languages: evidence from functional magnetic resonance imaging. *Neuroreport* 16, 761–765. doi: 10.1097/00001756-200505120-00021

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Garbin, G., Costa, A., Sanjuan, A., Forn, C., Rodriguez-Pujadas, A., Ventura, N., et al. (2011). Neural bases of language switching in high and early proficient bilinguals. *Brain Lang.* 119, 129–135. doi: 10.1016/j.bandl.2011.03.011

Gathercole, S. E., Pickering, S. J., Ambridge, B., and Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Dev. Psychol.* 40, 177–190. doi: 10.1037/0012-1649.40.2.177

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863. doi: 10.1038/13158

Giraud, A.-L., Truy, E., Frackowiak, R. S. J., Grégoire, M.-C., Pujol, J.-F., and Collet, L. (2000). Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. *Brain* 123, 1391–1402. doi: 10.1093/brain/123.7.1391

Gold, B. T., Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Snyder, A. Z., Salat, D. H., et al. (2005). Differing neuropsychological and neuroanatomical correlates of abnormal reading in early-stage semantic dementia and dementia of the Alzheimer type. *Neuropsychologia* 43, 833–846. doi: 10.1016/j.neuropsychologia.2004.10.005

Golestani, N., Alario, F., Meriaux, S., Le Bihan, D., Dehaene, S., and Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia* 44, 1029–1040. doi: 10.1016/j.neuropsychologia.2005.11.009

Gray, T., and Kiran, S. (2013). A theoretical account of lexical and semantic naming deficits in bilingual aphasia. *J. Speech Lang. Hear. Res.* 56, 1314–1327. doi: 10.1044/1092-4388(2012/12-0091)

Grodzinsky, Y., and Santi, A. (2008). The battle for Broca's region. *Trends Cogn. Sci.* 12, 474–480. doi: 10.1016/j.tics.2008.09.001

Grosjean, F. (1989). Neurolinguists beware! The bilingual is not two monolinguals in one person. *Brain Lang.* 36, 3–15. doi: 10.1016/0093-934x(89)90048-5

Grosjean, F., and Li, P. (2013). *The Psycholinguistics of Bilingualism.* Malden: Wiley-Blackwell.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., Hartwig, R., et al. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U S A* 107, 16494–16499. doi: 10.1073/pnas.1008121107

Hernandez, A. E., Hofmann, J., and Kotz, S. A. (2007). Age of acquisition modulates neural activity for both regular and irregular syntactic functions. *NeuroImage* 36, 912–923. doi: 10.1016/j.neuroimage.2007.02.055

Hernandez, A. E., Woods, E. A., and Bradley, K. A. L. (2015). Neural correlates of single word reading in bilingual children and adults. *Brain Lang.* 143, 11–19. doi: 10.1016/j.bandl.2015.01.010

Hesling, I., Dilharreguy, B., Bordessoules, M., and Allard, M. (2012). The neural processing of second language comprehension modulated by the degree of proficiency: a listening connected speech fMRI study. *Open Neuroimag. J.* 6, 44–54. doi: 10.2174/1874440001206010044

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hillis, A. E. (2001). "The organization of the lexical system," in *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*, ed. B. Rapp (New York, NY: Psychology Press), 185–210.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: which suggested differences can we trust and what do they mean? *Lang. Learn.* 56, 279–304. doi: 10.1111/j.1467-9922.2006.00365.x

Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Jamal, N. I., Piche, A. W., Napoliello, E. M., Perfetti, C. A., and Eden, G. F. (2012). Neural basis of single-word reading in Spanish-english bilinguals. *Hum. Brain Mapp.* 33, 235–245. doi: 10.1002/hbm.21208

Jeong, H., Hashizume, H., Sugiura, M., Sassa, Y., Yokoyama, S., Shiozaki, S., et al. (2011). Testing second language oral proficiency in direct and semidirect settings: a social-cognitive neuroscience perspective. *Lang. Learn.* 61, 675–699. doi: 10.1111/j.1467-9922.2011.00635.x

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., et al. (2007a). Effect of syntactic similarity on cortical activation during second language processing: a comparison of English and Japanese among native Korean trilinguals. *Hum. Brain Mapp.* 28, 194–204. doi: 10.1002/hbm.20269

Jeong, H., Sugiura, M., Sassa, Y., Yokoyama, S., Horie, K., Sato, S., et al. (2007b). Cross-linguistic influence on brain activation during second language processing: an fMRI study. *Biling. Lang. Cogn.* 10, 175–187. doi: 10.1017/s1366728907002921

Jeong, H., Sugiura, M., Suzuki, W., Sassa, Y., Hashizume, H., and Kawashima, R. (2015). Neural correlates of second-language communication and the effect of language anxiety. *Neuropsychologia* 66, 182–192. doi: 10.1016/j.neuropsychologia.2014.11.013

Jiang, M., Yang, L.-X., Jia, L., Shi, X., Wang, H., Wang, L.-Y., et al. (2015). Distinct activated cortical areas and volumes in Uygur-Chinese bilinguals. *Transl. Neurosci.* 6, 227–234. doi: 10.1515/tnsci-2015-0024

Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of english as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0

Kemeny, S., Xu, J., Park, G. H., Hosey, L. A., Wettig, C. M., and Braun, A. R. (2006). Temporal dissociation of early lexical access and articulation using a delayed naming task—an FMRI study. *Cereb. Cortex* 16, 587–595. doi: 10.1093/cercor/bhj006

Kim, C.-M. (2015). Neural bases underlying native or foreign word production, and language switching. *J. Korea Acad. Ind. Coop. Soc.* 16, 1707–1714. doi: 10.5762/kais.2015.16.3.1707

Kim, S. Y., Qi, T., Feng, X., Ding, J., Liu, L., and Cao, F. (2016). How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean-Chinese-English trilinguals. *Neuroimage* 129, 25–39. doi: 10.1016/j.neuroimage.2015.11.068

Kirschen, M. P., Chen, S. H. A., Schraedley-Desmond, P., and Desmond, J. E. (2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *Neuroimage* 24, 462–472. doi: 10.1016/j.neuroimage.2004.08.036

Klein, D., Milner, B., Zatorre, R. J., Zhao, V., and Nikelski, J. (1999). Cerebral organization in bilinguals: a PET study of Chinese-English verb generation. *Neuroreport* 10, 2841–2846. doi: 10.1097/00001756-199909090-00026

Koenigs, M., Barbey, A. K., Postle, B. R., and Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *J. Neurosci.* 29, 14980–14986. doi: 10.1523/jneurosci.3706-09.2009

Kovelman, I., Baker, S. A., and Petitto, L. A. (2008). Bilingual and monolingual brains compared: a functional magnetic resonance imaging investigation of syntactic processing and a possible "neural signature" of bilingualism. *J. Cogn. Neurosci.* 20, 153–169. doi: 10.1162/jocn.2008.20011

Koyama, M. S., Stein, J. F., Stoodley, C. J., and Hansen, P. C. (2013). Cerebral mechanisms for different second language writing systems. *Neuropsychologia* 51, 2261–2270. doi: 10.1016/j.neuropsychologia.2013.08.002

Kumar, U. (2014). Effect of orthography over neural regions in bilinguals: a view from neuroimaging. *Neurosci. Lett.* 580, 94–99. doi: 10.1016/j.neulet.2014.07.053

Kumar, U., Das, T., Bapi, R. S., Padakannaya, P., Joshi, R. M., and Singh, N. C. (2010). Reading different orthographies: an fMRI study of phrase reading in Hindi-English bilinguals. *Read. Writ.* 23, 239–255. doi: 10.1007/s11145-009-9176-8

Kuzmina, E., Goral, M., Norvik, M., and Weekes, B. (2019). What influences language impairment in bilingual aphasia? A meta-analytic review. *Front. Psychol.* 10:445. doi: 10.3389/fpsyg.2019.00445

Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., et al. (2009a). ALE meta-analysis workflows *via* the brainmap database: progress towards a probabilistic functional brain atlas. *Front. Neuroinform.* 3:23. doi: 10.3389/neuro.11.023.2009

Laird, A. R., Eickhoff, S. B., Li, K., Robin, D. A., Glahn, D. C., and Fox, P. T. (2009b). Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. *J. Neurosci.* 29, 14496–14505. doi: 10.1523/jneurosci.4004-09.2009

Lechtenberg, R., and Gilman, S. (1978). Speech disorders in cerebellar disease. *Ann. Neurol.* 3, 285–290. doi: 10.1002/ana.410030402

Lenneberg, E. H. (1967). *Biological Foundations of Language.* New York, NY: Wiley.

Li, Y., Yang, J., Scherf, K. S., and Li, P. (2013). Two faces, two languages: an fMRI study of bilingual picture naming. *Brain Lang.* 127, 452–462. doi: 10.1016/j.bandl.2013.09.005

Liu, H., and Cao, F. (2016). L1 and L2 processing in the bilingual brain: a meta-analysis of neuroimaging studies. *Brain Lang.* 159, 60–73. doi: 10.1016/j.bandl.2016.05.013

Liu, H., Hu, Z., Guo, T., and Peng, D. (2010). Speaking words in two languages with one brain: neural overlap and dissociation. *Brain Res.* 1316, 75–82. doi: 10.1016/j.brainres.2009.12.030

Locke, J. L., and Bogin, B. (2006). Language and life history: a new perspective on the development and evolution of human language. *Behav. Brain Sci.* 29, 259–280. doi: 10.1017/s0140525x0600906x

Long, M. H. (1990). Maturational constraints on language development. *Stud. Second Lang. Acquis.* 12, 251–285. doi: 10.1017/s0272263100009165

Luke, K. K., Liu, H. L., Wai, Y. Y., Wan, Y. L., and Tan, L. H. (2002). Functional anatomy of syntactic and semantic processing in language comprehension. *Hum. Brain Mapp.* 16, 133–145. doi: 10.1002/hbm.10029

Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: within- and between language competition. *Biling. Lang. Cogn.* 6, 97–115. doi: 10.1017/s1366728903001068

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C. H. S., Beaton, A., Desmond, J., et al. (2014). Consensus paper: language and the cerebellum: an ongoing enigma. *Cerebellum* 13, 386–410. doi: 10.1007/s12311-013-0540-5

McDonald, A. W., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the role of dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838. doi: 10.1126/science.288.5472.1835

Meschyan, G., and Hernandez, A. E. (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: an fMRI investigation. *Neuroimage* 29, 1135–1140. doi: 10.1016/j.neuroimage.2005.08.055

Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Murdoch, B. E., and Whelan, B.-M. (2007). Language disorders subsequent to left cerebellar lesions: a case for bilateral cerebellar involvement in language? *Folia Phoniatr. Logop.* 59, 184–189. doi: 10.1159/000102930

Nakada, T., Fujii, Y., and Kwee, I. L. (2001). Brain strategies for reading in the second language are determined by the first language. *Neurosci. Res.* 40, 351–358. doi: 10.1016/s0168-0102(01)00247-4

Nakagawa, H., Iwasaki, S., Kichikawa, K., Fukusumi, A., Taoka, T., Ohishi, H., et al. (1998). Normal myelination of anatomic nerve fiber bundles: MR analysis. *AJNR Am. J. Neuroradiol.* 19, 1129–1136.

Nelson, J. R., Liu, Y., Fiez, J., and Perfetti, C. A. (2009). Assimilation and accommodation patterns in ventral occipitotemporal cortex in learning a second writing system. *Hum. Brain Mapp.* 30, 810–820. doi: 10.1002/hbm.20551

Nosarti, C., Mechelli, A., Green, D. W., and Price, C. J. (2010). The impact of second language learning on semantic and nonsemantic first language reading. *Cereb. Cortex* 20, 315–327. doi: 10.1093/cercor/bhp101

Oh, A., Duerden, E. G., and Pang, E. W. (2014). The role of the insula in speech and language processing. *Brain Lang.* 135, 96–103. doi: 10.1016/j.bandl.2014.06.003

Palomar-García, M. Á., Bueichekú, E., Ávila, C., Sanjuán, A., Strijkers, K., Ventura-Campos, N., et al. (2015). Do bilinguals show neural differences with monolinguals when processing their native language? *Brain Lang.* 142, 36–44. doi: 10.1016/j.bandl.2015.01.004

Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2000). A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. *Neuroimage* 11, 347–357. doi: 10.1006/nimg.2000.0546

Paradis, M. (1994). "Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism," in *Implicit and Explicit Learning of Second Languages*, ed. N. Ellis (London: Academic Press), 393–419.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism.* Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages.* Amsterdam: John Benjamins.

Park, H. R., Badzakova-Trajkov, G., and Waldie, K. E. (2012). Language lateralization in late proficient bilinguals: a lexical decision fMRI study. *Neuropsychologia* 50, 688–695. doi: 10.1016/j.neuropsychologia.2012.01.005

Pasqualotto, A., Kobanbay, B., and Proulx, M. J. (2015). Neural stimulation has a long-term effect on foreign vocabulary acquisition. *Neural Plast.* 2015:671705. doi: 10.1155/2015/671705

Paulesu, E., Frith, C. D., and Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. doi: 10.1038/362342a0

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Hum. Brain Mapp.* 19, 170–182. doi: 10.1002/hbm.10110

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *Neuroreport* 7, 2439–2444. doi: 10.1097/00001756-199611040-00007

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain* 121, 1841–1852. doi: 10.1093/brain/121.10.1841

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., and Tan, L. H. (2007). Reading in two writing systems: accommodation and assimilation of the brain's reading network. *Biling. Lang. Cogn.* 10, 131–146. doi: 10.1017/s1366728907002891

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062

Price, C. J., and Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253. doi: 10.1016/j.tics.2011.04.001

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, R. T., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* 4, 8–26. doi: 10.1093/cercor/4.1.8

Reiterer, S. M., Hu, X., Suathi, T. A., and Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Front. Psychol.* 4:782. doi: 10.3389/fpsyg.2013.00782

Rodriguez-Fornells, A., Rotte, M., Heinze, H. J., Nösselt, T., and Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature* 415, 1026–1029. doi: 10.1038/4151026a

Román, P., Gonzáles, J., Ventura-Campos, N., Rodríguez-Pujadas, A., Sanjuán, A., and Ávila, C. (2015). Neural differences between monolinguals and early bilinguals in their native language during comprehension. *Brain Lang.* 150, 80–89. doi: 10.1016/j.bandl.2015.07.011

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modeling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage* 60, 830–846. doi: 10.1016/j.neuroimage.2011.11.050

Roux, F. E., Lubrano, V., Lauwers-Cances, V., Trémoulet, M., Mascott, C. R., and Démonet, J.-F. (2004). Intra-operative mapping of cortical areas involved in reading in mono- and bilingual patients. *Brain* 127, 1796–1810. doi: 10.1093/brain/awh204

Roux, F. E., and Trémoulet, M. (2002). Organization of language areas in bilingual patients: a cortical stimulation study. *J. Neurosurg.* 97, 857–864. doi: 10.3171/jns.2002.97.4.0857

Ruben, R. J. (1997). Time frame of critical/sensitive periods of language developments. *Acta Otolaryngol.* 117, 202–205. doi: 10.3109/00016489709117769

Samar, R. G., and Akbari, M. (2012). A language teacher in the haze of bilingual aphasia: a Kurdish-Persian case. *Procedia Soc. Behav. Sci.* 32, 252–257. doi: 10.1016/j.sbspro.2012.01.037

Saur, D., Baumgaertner, A., Moehring, A., Büchel, C., Bonnesen, M., Rose, M., et al. (2009). Word order processing in the bilingual brain. *Neuropsychologia* 47, 158–168. doi: 10.1016/j.neuropsychologia.2008.08.007

Sebastian, R., Laird, A. R., and Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Appl. Psycholinguist.* 32, 799–819. doi: 10.1017/s0142716411000075

Skeide, M. A., and Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* 17, 323–332. doi: 10.1038/nrn.2016.23

Suh, S., Yoon, H. W., Lee, S., Chung, J. Y., Cho, Z. H., and Park, H. (2007). Effects of syntactic complexity in L1 and L2; an fMRI study of Korean-English bilinguals. *Brain Res.* 1136, 178–189. doi: 10.1016/j.brainres.2006.12.043

Sun, Y., Peng, D., Ding, G., Qi, T., Desroches, A. S., and Liu, L. (2015). The dynamic nature of assimilation and accommodation procedures in the brains of Chinese-English and English-Chinese bilinguals. *Hum. Brain Mapp.* 36, 4144–4157. doi: 10.1002/hbm.22904

Talairach, J., and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging.* New York, NY: Thieme Medical Publishers, Inc.

Tan, L. H., Spinks, J. A., Feng, C. M., Siok, W. T., Perfetti, C. A., Xiong, J., et al. (2003). Neural systems of second language reading are shaped by native language. *Hum. Brain Mapp.* 18, 158–166. doi: 10.1002/hbm.10089

Tatsuno, Y., and Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency and task demands. *J. Neurosci.* 25, 1637–1644. doi: 10.1523/jneurosci.3978-04.2005

Titone, D., Libben, M., Mercier, J., Whitford, V., and Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: the effects of L2 knowledge, semantic constraint and L1–L2 intermixing. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1412–1431. doi: 10.1037/a0024492

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16, 765–780. doi: 10.1006/nimg.2002.1131

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *J. Psycholinguist. Res.* 30, 37–69. doi: 10.1023/A:1005204207369

Ullman, M. T. (2005). "A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model," in *Mind and Context in Adult Second Language Acquisition: Methods, Theory and Practice*, ed. C. Sanz (Washington: Georgetown University Press), 141–178.

Ullman, M. T. (2006). The declarative/procedural model and the shallow structure hypothesis. *Appl. Psycholinguist.* 27, 97–105. doi: 10.1017/s014271640606019x

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., and Hagoort, P. (2008). Language conflict in the bilingual brain. *Cereb. Cortex* 18, 2706–2716. doi: 10.1093/cercor/bhn030

Verreyt, N., Bogaerts, L., Cop, U., Bernolet, S., De Letter, M., Hemelsoet, D., et al. (2013). Syntactic priming in bilingual patients with parallel and differential aphasia. *Aphasiology* 27, 867–887. doi: 10.1080/02687038.2013.791918

Videsott, G., Herrnberger, B., Hoenig, K., Schilly, E., Grothe, J., Wiater, W., et al. (2010). Speaking in multiple languages: neural correlates of language proficiency in multilingual word production. *Brain Lang.* 113, 103–112. doi: 10.1016/j.bandl.2010.01.006

Vigneau, M., Beaucousin, V., Herve, P. Y., Jobard, G., Petit, L., Crivello, F., et al. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *Neuroimage* 54, 577–593. doi: 10.1016/j.neuroimage.2010.07.036

Vigneau, M., Jobard, G., Mazoyer, B., and Tzourio-Mazoyer, N. (2005). Word and non-word reading: What role for the visual word form area? *Neuroimage* 27, 694–705. doi: 10.1016/j.neuroimage.2005.04.038

Vingerhoets, G., Van Borsel, J., Tesink, C., van den Noort, M., Deblaere, K., Seurinck, R., et al. (2003). Multilingualism: an fMRI study. *Neuroimage* 20, 2181–2196. doi: 10.1016/j.neuroimage.2003.07.029

Wager, T. D., Ching-Yune, C. S., Lacey, S. C., Nee, D. E., Franklin, M., and Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *Neuroimage* 27, 323–340. doi: 10.1016/j.neuroimage.2005.01.054

Waldron, E. J., and Hernandez, A. E. (2013). The role of age of acquisition on past tense generation in Spanish-English bilinguals: an fMRI study. *Brain Lang.* 125, 28–37. doi: 10.1016/j.bandl.2013.01.002

Yang, J., Tan, L., and Li, P. (2011). Lexical representation of nouns and verbs in the late bilingual brain. *J. Neurolinguist.* 24, 674–682. doi: 10.1016/j.jneuroling.2011.07.002

Yokoyama, S., Okamoto, H., Miyamoto, T., Yoshimoto, K., Kim, J., Iwata, K., et al. (2006). Cortical activation in the processing of passive sentences in L1 and L2: An fMRI study. *Neuroimage* 30, 570–579. doi: 10.1016/j.neuroimage.2005.09.066

Yoon, H. W., Chung, J.-Y., Kim, K. H., Song, M.-S., and Park, H. W. (2006). An fMRI study of Chinese character reading and picture naming by native Korean speakers. *Neurosci. Lett.* 392, 90–95. doi: 10.1016/j.neulet.2005.09.027

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 12 July 2019
doi: 10.3389/fnhum.2019.00243

Top-Down Predictions of Familiarity and Congruency in Audio-Visual Speech Perception at Neural Level

Orsolya B. Kolozsvári[1,2]*, *Weiyong Xu*[1,2], *Paavo H. T. Leppänen*[1,2] *and* *Jarmo A. Hämäläinen*[1,2]

[1] *Department of Psychology, University of Jyväskylä, Jyväskylä, Finland,* [2] *Jyväskylä Centre for Interdisciplinary Brain Research (CIBR), University of Jyväskylä, Jyväskylä, Finland*

During speech perception, listeners rely on multimodal input and make use of both auditory and visual information. When presented with speech, for example syllables, the differences in brain responses to distinct stimuli are not, however, caused merely by the acoustic or visual features of the stimuli. The congruency of the auditory and visual information and the familiarity of a syllable, that is, whether it appears in the listener's native language or not, also modulates brain responses. We investigated how the congruency and familiarity of the presented stimuli affect brain responses to audio-visual (AV) speech in 12 adult Finnish native speakers and 12 adult Chinese native speakers. They watched videos of a Chinese speaker pronouncing syllables (/pa/, /pha/, /ta/, /tha/, /fa/) during a magnetoencephalography (MEG) measurement where only /pa/ and /ta/ were part of Finnish phonology while all the stimuli were part of Chinese phonology. The stimuli were presented in audio-visual (congruent or incongruent), audio only, or visual only conditions. The brain responses were examined in five time-windows: 75–125, 150–200, 200–300, 300–400, and 400–600 ms. We found significant differences for the congruency comparison in the fourth time-window (300–400 ms) in both sensor and source level analysis. Larger responses were observed for the incongruent stimuli than for the congruent stimuli. For the familiarity comparisons no significant differences were found. The results are in line with earlier studies reporting on the modulation of brain responses for audio-visual congruency around 250–500 ms. This suggests a much stronger process for the general detection of a mismatch between predictions based on lip movements and the auditory signal than for the top-down modulation of brain responses based on phonological information.

Keywords: speech perception, magnetoencephalography, audio-visual stimuli, audio-visual integration, familiarity

Edited by:
Xiaolin Zhou,
Peking University, China

Reviewed by:
Lihan Chen,
Peking University, China
Nai Ding,
Zhejiang University, China

Correspondence:
Orsolya B. Kolozsvári
orsolya.b.kolozsvari@jyu.fi

Received: *16 January 2019*
Accepted: *28 June 2019*
Published: *12 July 2019*

Citation:
Kolozsvári OB, Xu W,
Leppänen PHT and Hämäläinen JA
(2019) Top-Down Predictions
of Familiarity and Congruency
in Audio-Visual Speech Perception
at Neural Level.
Front. Hum. Neurosci. 13:243.
doi: 10.3389/fnhum.2019.00243

INTRODUCTION

In most cases speech perception relies on the seamless interaction and integration of auditory and visual information. Listeners need to efficiently process a rapid and complex stream of multisensory information, making use of both visual and auditory cues. We wanted to examine how lifelong exposure to audio-visual speech affects the brain mechanisms of cross-modal integration and mismatch. Auditory and visual cues can be presented either congruently or incongruently and this match or mismatch of features could be used to study the audio-visual processing of speech. Using magnetoencephalography (MEG), we studied how the effects of congruency and familiarity (i.e., whether the speech stimuli are part of the listener's phonology or not) of the auditory and visual features are reflected in brain activity.

Audio-visual speech has been shown to activate (in sequence) the sensory areas around 100 ms from stimulation onset in the auditory and visual cortices (Sams et al., 1991; Möttönen et al., 2004; Salmelin, 2007), then the superior temporal sulcus around 150 ms (Nishitani and Hari, 2002), which has been shown to play an important role in the perception and interpretation of movements (both facial and body) of the speaker (Puce et al., 1998; Iacoboni et al., 2001). The inferior parietal cortex has been shown to be activated at around 200 ms, which is suggested to be related to the connection of the STS to the inferior frontal lobe (Broca's area) (Nishitani and Hari, 2002) with stronger activations in the left hemisphere than in the right (Capek et al., 2004; Campbell, 2008). This is followed by activation in the frontal areas close to Broca's area around 250 ms (Nishitani and Hari, 2002).

It has been suggested (Campbell, 2008) that seeing speech can affect what is perceived in either a complementary or correlated way. In the complementary mode, vision offers further information about some aspects of speech, which are harder to detect only auditorily and which may depend on the clear visibility of the speaker's lower face. In the correlated mode, on the other hand, successful speech processing depends on the speech stream's temporal-spectral signature showing similar dynamic patterning across both the audible and visible channels.

Audio-visual mismatch is often examined from the point of view of congruency (Jones and Callan, 2003; Hein et al., 2007), where congruent and incongruent audio-visual pairs are contrasted. The assumption is that congruency should only have an effect on perception when the inputs of unimodal sources have been integrated (van Atteveldt et al., 2007). In terms of brain responses, the STS has been shown to be a critical brain area for multisensory integration and congruency of auditory and visual information in the case of both speech and non-speech stimuli. For example, Beauchamp et al. (2010) used TMS to disrupt brain activity in STS, while participants viewed audio-visual stimuli that have been shown to cause the McGurk effect (where incongruent auditory and visual speech cues presented together produce an illusory percept; McGurk and Macdonald, 1976). When TMS was applied to the left STS during the perception of McGurk pairs, the frequency of the McGurk percept was greatly reduced. This reduction, in the likelihood of the McGurk effect, demonstrates that the STS is an important cortical locus for the McGurk effect and plays an important part in auditory-visual integration in speech.

Furthermore, a broad network of brain regions in addition to the STS have been found in fMRI studies to show differences between brain responses to incongruent and congruent audio-visual speech, including the precentral gyrus (Jones and Callan, 2003), the inferior parietal lobule (Jones and Callan, 2003), the supramarginal gyrus (Jones and Callan, 2003), the superior frontal gyrus (Miller and D'Esposito, 2005), Heschl's gyrus (Miller and D'Esposito, 2005) and the middle temporal gyrus (Callan et al., 2004).

Previous studies examining audio-visual speech have found relatively early event-related brain potential (ERP) effects around N1 and P2 responses (Stekelenburg and Vroomen, 2007; Baart et al., 2014). In this case the visual information leads the auditory information, that is, lip movements can precede actual phonation for up to several hundreds of milliseconds (Stekelenburg and Vroomen, 2007). This visual information allows the observer to make predictions about several aspects of the auditory signal (e.g., content, timing). Studies have shown that the auditory-evoked N1 and P2 components of ERPs, at latencies of 100–150 and 200–250 ms, respectively, are attenuated and speeded up when the auditory signal is accompanied by visual speech (Klucharev et al., 2003; Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). This suggests early predictive effects of the visual information on the auditory stimulation. Furthermore, no attenuation in N1 was found when no visual anticipatory information about sound onset is present, indicating that the temporal information present in the visual stimulus, rather than the content of the sound, is key in audio-visual interaction (Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010).

However, the N1 and P2 responses seem to be sensitive to the stimulus material. This was shown by Baart et al. (2014), who investigated speech-specific audio-visual integration, where they used speech stimuli and sinewave speech, and found that N1 suppression occurs regardless of the type of stimuli, but P2 amplitude was only suppressed in relation to speech stimuli. They found congruency effects for responses to speech stimuli from around 200 ms after audio-visual incongruency became apparent, with ERPs being more negative for congruent stimuli than for incongruent stimuli. These early suppression effects were found when comparing the brain responses between the unimodal and the multimodal stimuli.

In addition, audio-visual speech congruency effects have also been reported in later time-windows. Arnal et al. (2009) investigated how the visual signal of an audio-visual stimulus affects auditory speech processing. In their experiment they recorded early visual and auditory responses to matching (congruent) and non-matching (incongruent) audio-visual syllables using MEG and found no effect of audio-visual incongruence in the early time-window (M100). They detected the earliest mismatch effect 120 ms after voice onset, followed by three more maxima at 250, 370, and 460 ms. Their findings indicated a multistep comparison between the top-down visual prediction and the bottom-up auditory signal.

Another aspect affecting audio-visual speech is the long-term memory representations of speech, that is, the familiarity of the speech itself. It has been documented that speech perception is altered by an individual's language experience. Iverson et al. (2003) found that listeners of different languages respond to distinct acoustic aspects of the same speech stimulus. They compared Japanese, German, and English speakers' responses to contrasts of /ra/ and /la/, where they had to rate whether the stimulus presented was a good exemplar of their own native-language phoneme. They found that American listeners attend to the third formant, which reliably distinguishes /r/ from /l/, while Japanese listeners attend more strongly to the second formant, which is critical for distinguishing Japanese phonemes, but is not at all helpful in distinguishing /r/ from /l/.

This and other studies suggest that the effects of language experience on speech perception are due to neural coding of the acoustic components that are critical to native-language processing (e.g., Kuhl, 2000, 2004). Such effects of language exposure are reflected in brain responses around 150–200 ms, for example in the modulation of the strength of the mismatch negativity (MMN), which is thought to tap into language-specific perceptual sensitivity (Näätänen et al., 1997, 2007; Winkler et al., 1999; Zhang et al., 2005, 2009). Language-specific phonetic-phonological analysis has been shown to start 100–200 ms following stimulus onset (Vihla et al., 2000; Näätänen et al., 2007). MMN or mismatch field (MMF) in EEG and MEG studies, respectively, have indicated access to phonological categories (Vihla et al., 2000; Näätänen et al., 2007) and the distinct processing of native and non-native phonetic contrasts (Näätänen et al., 1997, 2007) in this time-window.

By comparing two groups with different native languages (Finnish and Chinese), we aimed to see how long-term audio-visual representations affect speech perception by examining the congruency effects. Additionally, we aimed to distinguish the effects of familiarity, which is a learned aspect of speech, from congruency, which should be an inherent aspect of the audio-visual stimuli related to the general correspondence between mouth movements and speech signal.

To this end, we compared brain responses measured with MEG to unfamiliar and familiar (called aspirated and unaspirated, respectively, see section "Materials and Methods" below) and also congruent and incongruent audio-visual speech stimuli. We expected to find significant differences in responses to congruent and incongruent stimuli for both Chinese and Finnish participants with larger responses to incongruent stimuli starting from 150 ms or later based on the previous literature (e.g., Arnal et al., 2009). However, in the case of the Finnish participants, we expected differences between the familiar and unfamiliar stimuli specifically starting in the same time-window as the congruency effect (150 ms onward), with the unfamiliar stimuli producing a larger response than the familiar stimuli if long-term phonological representations facilitate the processing of audio-visual speech.

MATERIALS AND METHODS

Participants

Participants were adult Finnish native speakers and adult Chinese native speakers studying in Jyväskylä, Finland. None of the participants had neurological or learning problems, hearing difficulties, using medication affecting the central nervous system, head injuries, ADHD or language-specific disorders. They all had normal or corrected-to-normal sight. The Finnish participants had no exposure to the Chinese language. In total, 19 Finnish native speakers and 18 Chinese native speakers participated in the study. Of these, 13 were excluded from the analysis due to excessive head movement (two participants), poor vision after correction (two participants), technical problems during recording (three participants), strong noise interference (two participants), or otherwise bad signal quality (four participants). Data included in the analysis were from 12 Finnish participants and 12 Chinese participants (see **Table 1** for characteristics of participants included).

Ethical approval for the study was provided by the Ethical Committee of the University of Jyväskylä. Participants gave their written informed consent to participate in the study. All participants received movie tickets as compensation for participating in the study.

Stimuli

The stimuli were video recordings of the syllables /pa/, /pha/, /ta/, /tha/ and /fa/. Of these five syllables, /fa/ was used for a cover task to maintain participants' attention on the stimuli [see **Figure 1** for oscillograms, spectrograms and acoustic features of the stimuli. Figures were created using Praat (Boersma and Weenink, 2018), see **Table 2** for description of the stimuli]. The videos were recorded using a Canon Legria HF200 HD video camera and were edited in Adobe Premier Pro CS5.5 to be 1800 ms long. The videos were recordings of a male native Mandarin Chinese speaker.

For the Finnish participants, /pa/ and /ta/ were considered familiar stimuli because they are part of their native phonology. For the Chinese participants all four syllables were familiar. The recordings could be audio only, in which the participant was presented with the audio track and the still image of the speaker; visual only, in which the video was presented without any sound; and audio-visual, where both audio track and video were presented at the same time. The audio-visual condition could be congruent, where what they saw was what they heard, or incongruent, where the audio did not match the video.

Procedure

Participants sat in a magnetically shielded, sound-attenuated room. They sat under the MEG helmet in a 68° sitting position.

Stimuli were presented using Presentation software (version 18.1; Neurobehavioral Systems, Inc., Albany, CA, United States) running on a Microsoft Windows computer using a Sound Blaster Audigy RX sound card and NVIDIA Quadro K5200 video card.

The stimuli were presented on a projector screen. Stimuli were projected from outside of the measurement room onto a mirror then reflected onto the projector screen using a Barco

TABLE 1 | Participant characteristics.

Native language	Finnish	Chinese
Mean age (*SD*)	23.92 (1.98)	24.75 (3.39)
Gender ratio (male:female)	6:6	3:9
Handedness ratio (right:left)	12:0	12:0

Mean age, gender ratio and handedness are for those included in the analysis.

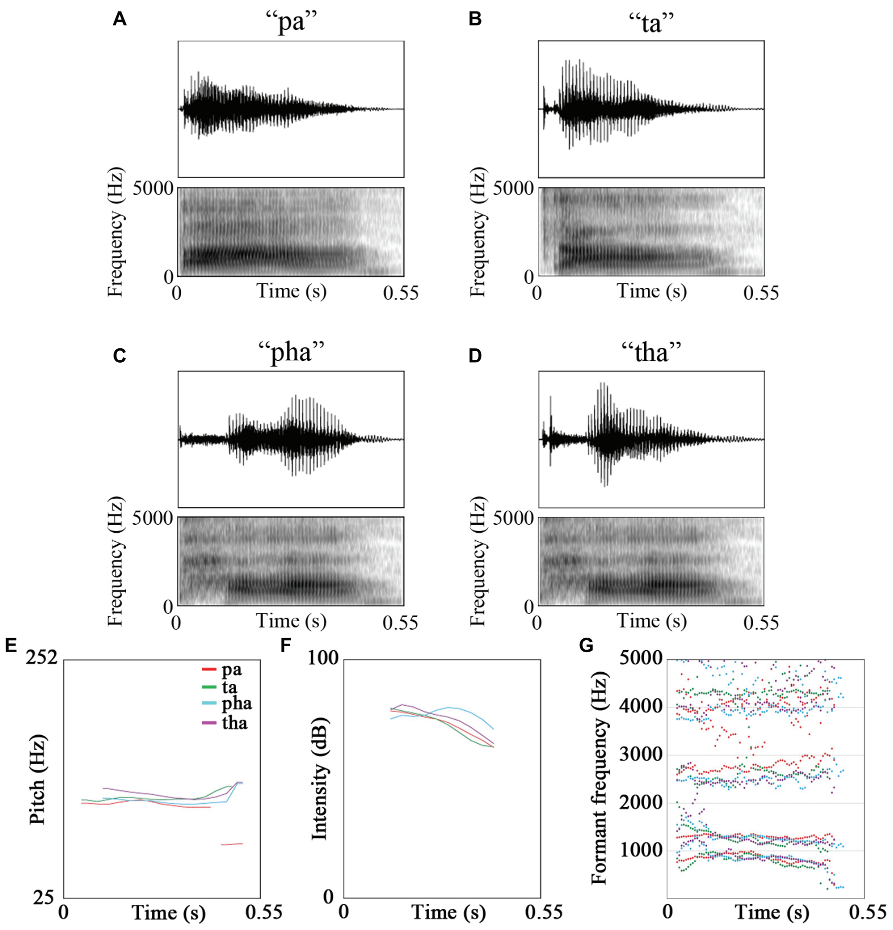


FIGURE 1 | Oscillograms, spectrograms and the acoustic features of the stimuli **(A)** pa, **(B)** ta, **(C)** pha, **(D)** tha, **(E)** pitch, **(F)** intensity, **(G)** formant frequencies (red - /pa/, green - /ta/, cyan - /pha/, purple - /tha/).

TABLE 2 | Stimuli description.

Modality	Target	Familiar / Unaspirated		Unfamiliar / Aspirated	
Audio	fa A	pa A	ta A	pha A	tha A
Visual	fa V	pa V	ta V	pha V	tha V
AV congruent	fa V / fa A	pa V / pa A	ta V/ ta A	pha V /pha A	tha V / tha A
AV incongruent	–	pa V / tha A	ta V / pha A	pha V / ta A	tha V / pa A

FL35 projector. The participants were sitting 1 m from the projection screen.

The participants were asked to watch short videos of a speaker uttering syllables and to attend to all stimuli presented. The videos were cropped to the mouth area of the speaker (from just above the nose to the clavicles). The fixation cross before the onset of the video clip was centered on where the lips of the speaker were in the videos. Videos were presented on a black background, in the center of the screen. The lights were dimmed. Sounds were presented through insert earphones (Lo-Fi auditory stimulation system, Elekta MEGIN Triux) at ∼70 dB sound pressure level.

The participants were presented with a blank screen for 500 ms, then a fixation cross for 550 ms, followed by a still image of the speaker for 500 ms and finally the stimuli, which was 1800 ms long.

The participants received eight practice trials. In the actual experiment 220 stimuli (20 targets for the cover task, and 50 audio-visual congruent, 50 audio-visual incongruent, 50 audio and 50 visual stimuli; /pa/ and /ta/ repeated 12 times each, /pha/ and /tha/ repeated 13 times each) were presented in pseudo-random order with no immediate repetitions of the same stimuli. Stimuli were presented in two blocks, with a short break (duration determined by the participant) in between the blocks (see **Figure 2** for a schematic representation of the video sequence and timings).

As a cover task the participants were asked to press a button to indicate if they saw and/or heard the target syllable /fa/.

Magnetoencephalography Recording and Preprocessing

The MEG data were recorded by a whole-head 306 channel Elekta Neuromag TRIUX MEG device in Jyväskylä, Finland, including

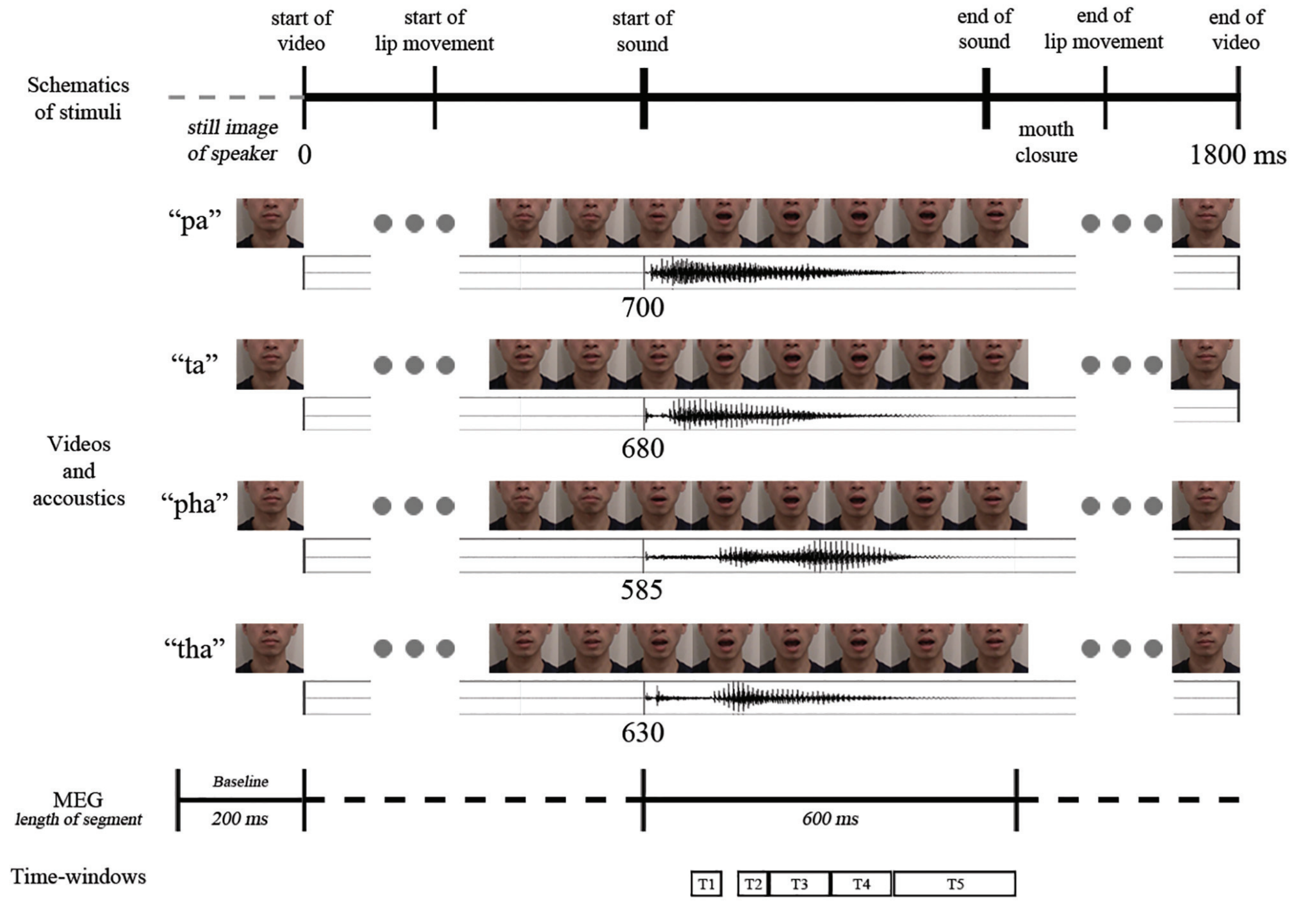


FIGURE 2 | Schematic representation of the temporal structure of the four congruent audio-visual stimuli used (above: video frames, below: oscillogram) and the analysis intervals (marked with T1–T5).

102 magnetometers and 204 orthogonal planar gradiometers. EOG was measured from two diagonally placed electrodes, slightly above the right eye and slightly below the left eye, with the ground electrode on the right clavicle. Five head position indicator (HPI) coils were attached to the scalp, three on the forehead and one behind each ear, and were used to monitor the location of the head in relation to the sensors during the recording by sending 293, 307, 314, 321, and 328 Hz sinusoidal currents into the five coils, respectively. The Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, United States) was used to determine the position of the HPI coils in relation to three anatomical landmarks (the nasion, left and right preauricular points). For co-registration purposes an additional set of scalp points (>100) were also digitized, distributed randomly over the skull.

Magnetoencephalography data were collected with a sampling rate of 1000 Hz and an online filter of 0.1–330 Hz. All data were preprocessed using the temporal extension of the signal space separation (tSSS) method with buffers of 30 s (Taulu and Kajola, 2005; Taulu et al., 2005) in Maxfilter 3.0^{TM} (Elekta AB) to remove external interference and correct for head movements. Bad channels were identified by visual inspection and marked for exclusion and reconstructed by the MaxFilter program. Head position was estimated in 200 ms time-windows and 10 ms steps for movement compensation.

Data were preprocessed using MNE Python (0.16.2) (Gramfort et al., 2013). Independent component analysis (ICA) using the fastICA algorithm (Hyvärinen and Oja, 2000) was applied to remove eye blinks, horizontal eye movements and cardiac artifacts. Data were low-pass filtered at 35 Hz using a zero-phase FIR filter with a bandwidth of 8.8 Hz. Then the continuous MEG recording was epoched into 200 ms before to 1800 ms after the onset of the video stimuli in the audio-visual condition. The epoched data were baselined using the 200 ms preceding the onset of stimuli. The epochs were shortened and realigned to 200 ms before and 1000 ms after the start of sound in the audio-visual condition. Data were then manually checked to remove any head movement–related artifacts and electronic jump artifacts. MEG epochs exceeding 2 pT/cm for gradiometer or 4 pT for magnetometer peak-to-peak amplitudes were excluded from further analysis. After artifact rejection, an average of 96.50% of trials were used for analysis. Event-related fields were obtained by averaging trials for different conditions separately. The data were then resampled to 250 Hz to shorten the computation time in the statistical analysis.

Statistical analysis of sensor-level data was done in FieldTrip toolbox (downloaded 20 October 2016; Oostenveld et al., 2011) for MATLAB R2016b (The MathWorks Inc., Natick, MA, 2000) while source-level analyses were run in MNE Python.

Time-Windows

Based on previous literature, five time-windows were investigated: 75–125, 150–200, 200–300, 300–400, and 400–600 ms (where 0 ms is the start of the sound in the section "Stimuli" as described above). The first time-window encompasses the basic auditory N1 m response (Poeppel et al., 1996; Parviainen et al., 2005; Salmelin, 2007), where the brain extracts speech sounds and their sequences from the incoming auditory signal and the responses are expected to be in the auditory cortices. The second time-window has been shown to be involved in further phonemic processing of the stimulus (Näätänen et al., 1997, 2007; Salmelin, 2007) with responses localized to the temporal cortex. The third time-window has been shown to be responsive to lexical-semantic manipulations (Helenius et al., 2002; Kujala et al., 2004) as well as to audio-visual manipulations (e.g., Raij et al., 2000; Arnal et al., 2009, around 250 ms), as have the fourth (Arnal et al., 2009, around 370 ms; Baart et al., 2014, 300–500 ms after onset of AV congruency) and the fifth time-windows (Arnal et al., 2009, around 460 ms).

Sensor-Level Analysis

Averaged planar gradiometer data were transformed into combined planar gradients using the vector sum of the two orthogonal sensors at each position implemented in the Fieldtrip toolbox (Oostenveld et al., 2011), which were then used for sensor-level analysis. Gradiometers were chosen because they are less sensitive to noise sources originating far from the sensors than magnetometers are.

Permutation tests with spatial and temporal clustering based on t-test statistics were carried out for planar gradients of individual averaged ERFs (Maris and Oostenveld, 2007). The five time-windows defined (see above) were investigated separately, with a cluster α level of 0.05 and the number of permutations 3000.

Source-Level Analysis

Source analysis was carried out with a minimum-norm estimate on the event-related fields of the magnetometers and gradiometers (Hämäläinen and Ilmoniemi, 1994). The noise covariance matrix was calculated from the baseline period of 200 ms preceding the start of the video (i.e., the participants were viewing the still image of the speaker).

Individual magnetic resonance images (MRI) were not available from the participants and therefore Freesurfer (RRID:SCR_001847) average brain (FSAverage) was used as a template for the source analysis (see below). Three-parameter scaling was used to co-register FSAverage with individual digitized head points. The average co-registration error was 3.54 mm (*SD*=0.27). A single layer BEM (Boundary Element Method) solution was used for the forward modeling.

Depth-weighted L2-minimum-norm estimate (wMNE) (Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006) was calculated for 4098 current dipoles with free orientation distributed on the cortical surface in each hemisphere. Dynamic statistical parametric mapping (dSPM) (Dale et al., 2000) was used to noise-normalize the inverse solution for further statistical analysis. Cluster-based permutation statistics in MNE Python were run on the dSPM source waveforms.

Statistical Analyses

Accuracy and reaction times in the cover task were examined using Target type (Audio only, Visual only, Audio-Visual) by Native language (Finnish, Chinese) ANOVAs.

Congruency and familiarity effects were examined using the interaction of Stimulus by Native language by comparing difference waves between the groups. If no significant results were obtained, Stimulus main effects were investigated between the stimuli. For comparisons investigating congruency, we compared responses to the congruent and incongruent audio-visual stimuli. For comparisons investigating familiarity, we compared responses to the congruent unaspirated audio-visual (/pa/ and /ta/ syllables) and the congruent aspirated audio-visual (/pha/ and /tha/ syllables) stimuli.

RESULTS

Behavioral Performance

Participants' accuracy scores were close to 100% (Finnish: 97.88%; Chinese: 98.35%) (**Table 3**), indicating that they were indeed paying attention to the stimuli. Accuracy (percentage of correct responses) were averaged for each participant, and a 3 (Target type: Audio only, Visual only, Audio-Visual) × 2 (Native language: Finnish, Chinese) repeated measures ANOVA resulted in no significant interaction or main effects.

Reaction times were on average 1189.72 ms (*SD*: 125.86) (**Table 4**). Reaction times were averaged for each participant, and a 3 (Target type: Audio only, Visual only, Audio-Visual) × 2 (Native language: Finnish, Chinese) repeated measures mixed ANOVA resulted in a significant Target type main effect [$F(1.954,42.985) = 6.338$, $p = 0.004$, partial $\eta 2 = 0.224$]. *Post hoc* t tests revealed that there was a significant difference between response time to visual only and audio only targets [$t(23) = 2.943$, $p = 0.007$], and audio-visual and audio only targets [$t(23) = 3.518$, $p = 0.002$] with audio only targets having longer reaction times than the other targets.

TABLE 3 | Accuracy scores for the Finnish and Chinese participants in detecting the target syllable /fa/.

	Accuracy (% of correct response to the target stimulus)			
	AV stimuli (%)	A stimuli (%)	V stimuli (%)	All stimuli (%)
Finnish (n = 12)	100	97.22	96.43	97.88
Chinese (n = 12)	98.81	98.61	97.62	98.35
Total (n = 24)	99.40	97.92	97.02	98.12

TABLE 4 | Reaction times for the Finnish and Chinese participants in detecting the syllable /fa/.

	Reaction times in ms (*SD*)			
	AV stimuli	**A stimuli**	**V stimuli**	**All stimuli**
Finnish (*n* = 12)	1170.56 (94.06)	1230.43 (94.51)	1187.56 (141.20)	1193.69 (103.84)
Chinese (*n* = 12)	1152.81 (151.20)	1201.29 (83.87)	1142.48 (160.23)	1163.16 (127.68)
Total (*n* = 24)	1161.69 (123.48)	1215.86 (88.64)	1165.02 (149.48)	1178.42 (114.88)

MEG

Our focus was on the native language interactions and we first examine, and report results with significant native language effects. In the absence of interactions, we report the main effects of congruency and familiarity.

Grand average plots of responses at sensor and source level for the congruency comparison and the familiarity comparison can be seen in **Supplementary Figures S1**, **S2**, respectively.

Sensor-Level Analysis

Congruency Effects

No significant effects were found in the first, second, third or fifth time-windows.

In the fourth time-window, two clusters were found to be significant for the Congruency main effect (responses to the incongruent stimuli compared to responses to the congruent stimuli) after the cluster permutation tests. One cluster (p = 0.036654) was found in the left frontal areas and another cluster (p = 0.046651) was found in the right temporal areas. See **Figure 3** for the topographic maps of brain responses in this time-window. See **Figure 4** for the topographic maps of the clusters and the average evoked responses from the channels forming the clusters.

Familiarity Comparison (Audio-Visual)

No significant statistical effects were found in the five time-windows examined using the cluster permutation tests.

Source-Level Analysis

Congruency Effects

No significant differences were found in the first, second, third and fifth time-windows.

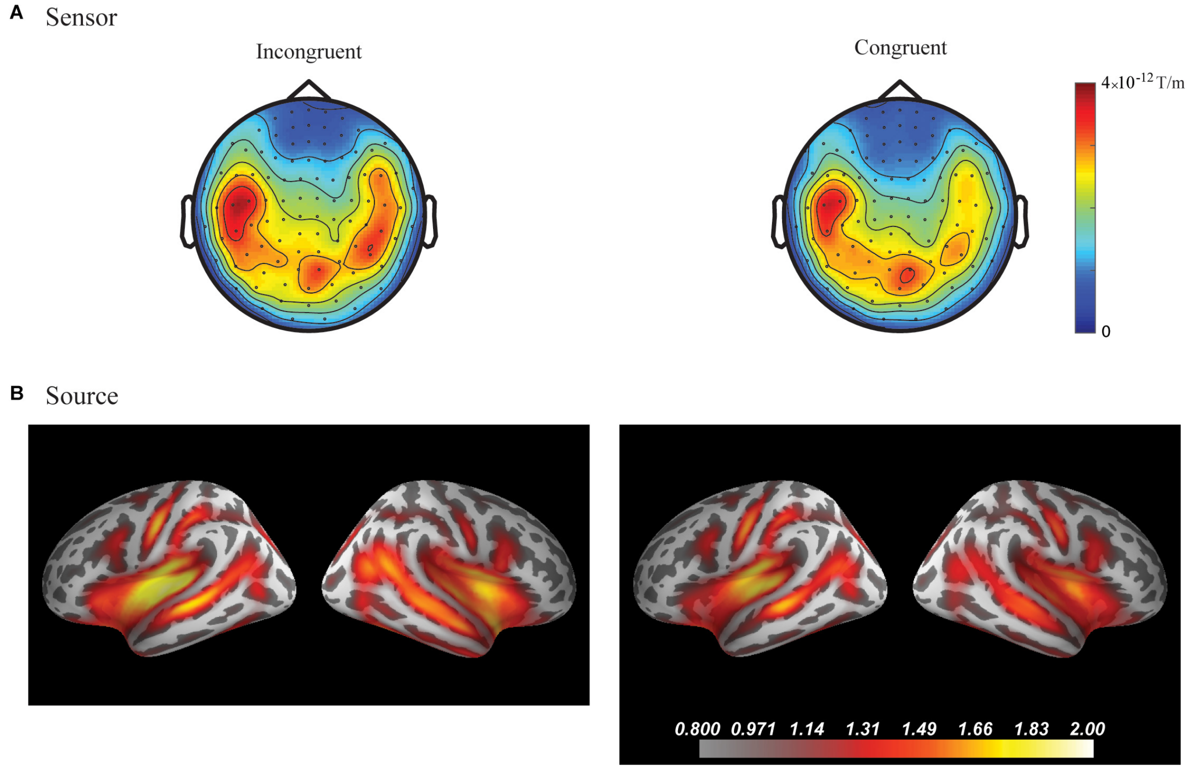


FIGURE 3 | Grand average plots at sensor and source level for incongruent and congruent audio-visual stimuli in the fourth time-window (300–400 ms) for the combined group (Chinese and Finnish speakers, *N* = 24). **(A)** Magnetic field topography of grand average evoked responses from combined planar gradients in the fourth time-window. **(B)** Dynamic statistical parametric maps (dSPM) of source activation of the grand average evoked responses in the fourth time-window.

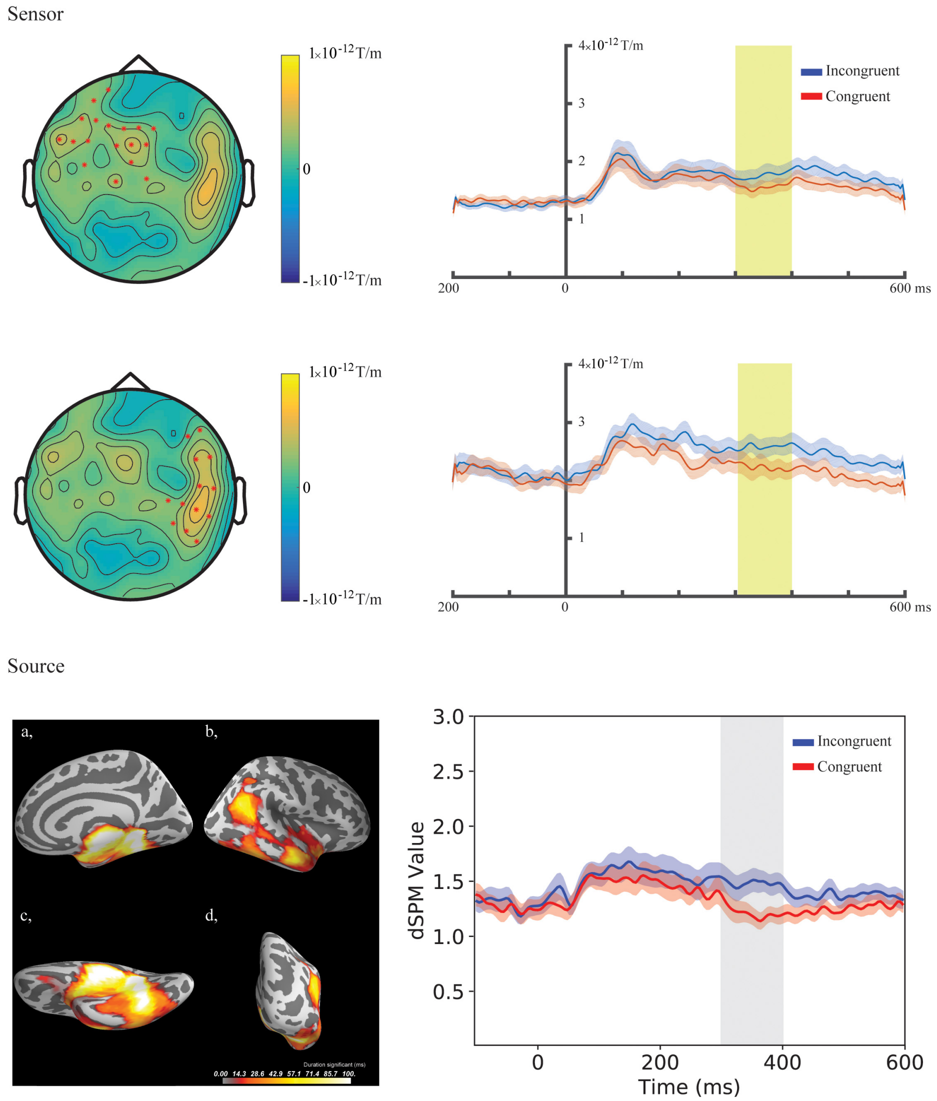


FIGURE 4 | Sensor- and source-level clusters based on the permutation statistics of the congruency effects for the combined group (Chinese and Finnish speakers, $N = 24$). **Left:** Clusters are represented by red dots in the sensor space and yellow and red colouring on the cortical surfaces for the source space. The brightness of the cluster was scaled by the temporal duration of the cluster in the source space. Lower left corner shows the cluster-based permutation statistics results for the incongruent vs congruent comparison in the source space: (a) Medial view, (b) Lateral view, (c) Ventral view, (d) Caudal view. **Right:** average evoked responses from the channels forming the cluster for the sensor space results and the source waveform (dSPM value) extracted from the clusters for the source space results. The red and blue shaded area represents the standard error of the mean and the yellow and gray shaded area indicates the time-window of the cluster.

In the fourth time-window, one cluster was found to be significant ($p = 0.039$) after the cluster permutation tests for the Congruency main effect (responses to the incongruent stimuli compared to responses to the congruent stimuli). The cluster encompassed the right temporal-parietal and medial areas. See **Figure 3** for dynamic statistical parametric maps (dSPM) source activation in this time-window. See **Figure 4** for the source waveform (dSPM value) extracted from the significant cluster.

Familiarity Comparison (Audio-Visual)

No significant statistical effects were found in the five time-windows examined using the cluster permutation tests.

All non-significant results of the permutation tests in the five time-windows, with lowest *p*-values, are reported in the **Supplementary Material 3**.

DISCUSSION

We investigated how the congruency and familiarity of a stimulus could affect audio-visual speech perception in two groups of adults, native speakers of Chinese and those of Finnish. The Chinese participants had long-term exposure to all of the stimuli because they belonged to their native language, but some of the speech sounds were not part of Finnish phonology, thus making them unfamiliar for the Finnish participants. We found significant differences in the congruency comparisons across these groups. A significant congruency main effect was found in the frontal and temporal regions at the sensor level and in the right temporal-parietal regions at the source level 300–400 ms following the onset of sound, but no significant effects were found for familiarity comparisons. Matching and mismatching audio-visual speech thus produces robust and replicable processing differences in the brain, which is consistent with findings in earlier studies. Direct comparison of responses to stimuli familiar (unaspirated) and unfamiliar (aspirated) to the Finnish participants do not show evidence for strong cross-modal top-down predictions that would modulate obligatory sensory brain responses.

We found a significant difference between the responses to the congruent and incongruent stimuli for Chinese and Finnish participants in the 300–400 ms time-window bilaterally at the sensor level at the left frontal and right temporal areas as well as in the right hemisphere at the source level in the temporal-parietal areas, indicating that both groups detected the incongruency. The time-window is in line with similar earlier studies using native language stimuli where the incongruence effects were found around 300–500 ms (Arnal et al., 2009; Baart et al., 2014). The localization of the congruency effect seems to depend on the task and contrast used. For example, left hemisphere emphasis was found using more complex stimulation with six different syllables (Arnal et al., 2009) and left frontotemporal regions for symbol–speech sound comparisons (Xu et al., 2019).

The direction of the congruency effect was also in line with earlier studies using audio-visual stimuli showing more brain activity for the incongruent compared to the congruent stimuli (e.g., Arnal et al., 2009; Xu et al., 2019). The direction of the effect likely indicates the benefit of using two modalities to decode the speech signal reflected in less allocation of neuronal resources to the process when the two modalities match (e.g., Bernstein and Liebenthal, 2014). For the incongruent stimuli, the brain response likely includes an error detection signal for the mismatching auditory and visual input. Similar to Arnal et al. (2009), we compared responses to congruent and incongruent stimuli. In their study, they found significant differences in relatively late time-windows, which showed multiple steps for audio-visual processing (with differences at ∼250, ∼370, and ∼460 ms, with responses being larger for the congruent stimuli at the first time-point, and larger for the incongruent stimuli at the later time-points) localized to the auditory cortex and the STS.

The lack of congruency effects in the time-windows after 400 ms in this study could be due to the differences in the complexity of the experimental design used, the features of the stimulus material and the timing parameters between the auditory and visual features of the present study and earlier studies. For example, in Arnal et al. (2009) audio-visual combinations of five different syllables were used, which made the identification of congruency more difficult and possibly required further processing steps compared to the current study.

Furthermore, we found no early effects of congruency at N1 m response (75–25 ms following sound onset), which is in line with previous observations (Stekelenburg and Vroomen, 2007). Our results corroborate the assumption that early responses are predominantly sensitive to the stimulus material used for the comparisons. Differences found in the N1 and P2 time-windows were related to suppression effects of audio-visual stimuli compared to audio only stimuli, and not to the direct comparison of congruent and incongruent audio-visual stimuli (van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007).

The source localization result of the current study was in line with the sensor-level results in terms of the time-window. However, the clusters at the source level were observed only in the right hemisphere and in a widely spread area encompassing the superior temporal areas as well as the medial and ventral surfaces of the temporal lobe. The superior temporal cortex is roughly in line with that found in Arnal et al. (2009). The widely spread clustering at the source level could be due to methodological limitations. It is important to note that we used a template MRI, and this could have increased the localization error of the brain responses in the source-level analysis. Furthermore, the difference was found in a relatively late time-window and appears quite widespread in time, and the localization of ongoing activation can be more challenging than those of clear time-locked evoked responses. These might explain the differences in the locations of the clusters between the sensor and source level, although we assume they reflect the same effect.

We found no significant effects of familiarity when directly comparing the responses to stimuli that were part of the participants' native language and to stimuli that were not part of their native language. The earlier studies have mostly examined this in auditory oddball experiments investigating deviance detection based on categorical perception of phonemes (e.g., Näätänen et al., 1997; Winkler et al., 1999). First, having equal probabilities of presentation for each stimulus type allows examination of the obligatory sensory responses without overlap from other processes. However, our null results comparing the responses to these stimuli in a passive cross-modal task suggest that the use of either an active comparison involving phonological representations or an identification task which would actively engage these representations is needed to lead to differences in brain activity for familiar and unfamiliar speech stimuli. Second, we examined evoked responses to audio-visual stimuli instead of induced brain activity. It is possible that the familiarity effects could produce brain activity that is not phase-locked to the stimuli. In this case the effect would not be observable in evoked responses. However, we did not have a hypothesis on the specific frequency band or time-window, where the difference in induced activity could be observed. Future studies could examine this in more detail.

The familiarity of speech in our study referred to whether participants perceiving the stimuli had prior knowledge of them, i.e., whether the syllables were present in their native phonology or not. Our stimuli (syllables) were produced by a native Chinese, non-finish speaker. This was required as native Finnish speakers would not be able to naturally produce all stimuli used in the experiment. Future studies could examine the effect of the speaker identity by using recordings of both native Chinese speaker and native Finnish speaker and how it might interact with the phonological familiarity of speech sounds.

CONCLUSION

Our results show that in the case of audio-visual speech stimuli, congruency has an effect around 300 to 400 ms after the start of voicing. This effect was found in the temporal-parietal brain areas, partly replicating earlier findings. We found no significant differences between Chinese and Finnish speakers in their brain responses depending on the familiarity of the speech stimuli, that is, whether the syllables belonged to the native language or not. This suggests that the congruency effect is a result of a general detection of a mismatch between prediction based on lip movements and the auditory signal.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Ethics Committee of the University of Jyväskylä with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the University of Jyväskylä.

AUTHOR CONTRIBUTIONS

OK, JH, and WX designed the study, performed the MEG experiments, and analyzed the data. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by the European Union Projects ChildBrain (Marie Curie Innovative Training Networks, # 641652), Predictable (Marie Curie Innovative Training Networks, # 641858), and the Academy of Finland (MultiLeTe #292 466).

ACKNOWLEDGMENTS

The authors would like to thank Chunhan Chiang for his help with data collection and constructive discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00243/full#supplementary-material

FIGURE S1 | Grand average plots at sensor and source level for incongruent and congruent audio-visual stimuli for the two groups. **(a)** Grand averaged waveform for the combined planar gradient (vector sum of the paired orthogonal gradiometer channels) channels grouped (channels included indicated by circles) over the left and right temporal channels in the Chinese (above, *N*=12) and Finnish (below, *N*=12) groups. **(b)** Magnetic field topography and dynamic statistical parametric maps (dSPM) source activation of the grand average evoked responses in the five time-windows investigated in the study (75–125, 150–200, 200–300, 300–400, and 400–600 ms) for the two conditions.

FIGURE S2 | Grand average plots at sensor and source level for unfamiliar and familiar congruent audio-visual stimuli for the two groups. **(a)** Grand averaged waveform for the combined planar gradient (vector sum of the paired orthogonal gradiometer channels) channels grouped (channels included indicated by circles) over the left and right temporal channels in the Chinese (above, *N*=12) and Finnish (below, *N*=12) groups. **(b)** Magnetic field topography and dynamic statistical parametric maps (dSPM) source activation of the grand average evoked responses in the five time-windows investigated in the study (75–125, 150–200, 200–300, 300–400, and 400–600 ms) for the two conditions.

REFERENCES

Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29, 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009

Baart, M., Stekelenburg, J. J., and Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia* 53, 115–121. doi: 10.1016/j.neuropsychologia.2013.11.011

Beauchamp, M. S., Nath, A. R., and Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the mcgurk effect. *J. Neurosci.* 30, 2414–2417. doi: 10.1523/JNEUROSCI.4865-09.2010

Bernstein, L. E., and Liebenthal, E. (2014). Neural pathways for visual speech perception. *Front. Neurosci.* 8:386. doi: 10.3389/fnins.2014.00386

Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234. doi: 10.1111/j.1460-9568.2004.03670.x

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer [Computer Program] (Version 6.0.37).* Available at: http://www.praat.org/ (accssed July 07, 2017).

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., and Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J. Cogn. Neurosci.* 16, 805–816. doi: 10.1162/089892904970771

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 1001–1010. doi: 10.1098/rstb.2007.2155

Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., and Neville, H. J. (2004). The cortical organization of audio-visual sentence comprehension: an fMRI study at 4 tesla. *Cogn. Brain Res.* 20, 111–119. doi: 10.1016/j.cogbrainres.2003.10.014

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R., Belliveau, J. W., and Lewine, J. D. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/BF02512476

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/JNEUROSCI.1740-07.2007

Helenius, P., Salmelin, R., Connolly, J. F., Leinonen, S., and Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J. Neurosci.* 22, 2936–2944. doi: 10.1523/JNEUROSCI.22-07-02936.2002

Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.-C., et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13995–13999. doi: 10.1073/pnas.241474598

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57. doi: 10.1016/S0010-0277(02)00198-1

Jones, J. A., and Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14, 1129–1133. doi: 10.1097/00001756-200306110-00006

Klucharev, V., Möttönen, R., and Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cogn. Brain Res.* 18, 65–75. doi: 10.1016/j.cogbrainres.2003.09.004

Kuhl, P. K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11850–11857. doi: 10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Cogn. Brain Res.* 21, 106–113. doi: 10.1016/j.cogbrainres.2004.05.011

Lin, F. H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31, 160–171. doi: 10.1016/j.neuroimage.2005.11.054

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

McGurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Miller, L. M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893. doi: 10.1523/JNEUROSCI.0896-05.2005

Möttönen, R., Schürmann, M., and Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neurosci. Lett.* 363, 112–115. doi: 10.1016/j.neulet.2004.03.076

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nishitani, N., and Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron* 36, 1211–1220. doi: 10.1016/S0896-6273(02)01089-9

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Parviainen, T., Helenius, P., and Salmelin, R. (2005). Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. *Cereb. Cortex* 15, 1054–1063. doi: 10.1093/cercor/bhh206

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P. L., Rowley, H. A., Wexler, K., et al. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cogn. Brain Res.* 4, 231–242. doi: 10.1016/S0926-6410(96)00643-X

Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* 18, 2188–2199. doi: 10.1523/JNEUROSCI.18-06-02188.1998

Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/S0896-6273(00)00138-0

Salmelin, R. (2007). Clinical neurophysiology of language: the MEG approach. *Clin. Neurophysiol.* 118, 237–254. doi: 10.1016/j.clinph.2006.07.316

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., et al. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.* 127, 141–145. doi: 10.1016/0304-3940(91)90914-F

Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973. doi: 10.1162/jocn.2007.19.12.1964

Taulu, S., and Kajola, M. (2005). Presentation of electromagnetic multichannel data: the signal space separation method. *J. Appl. Phys.* 97:124905. doi: 10.1063/1.1935742

Taulu, S., Simola, J., and Kajola, M. (2005). Applications of the signal space separation method. *IEEE Trans. Signal Process.* 53, 3359–3372. doi: 10.1109/TSP.2005.853302

van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007). Top–down task effects overrule automatic multisensory responses to letter–sound pairs in auditory association cortex. *NeuroImage* 36, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102

Vihla, M., Lounasmaa, O. V., and Salmelin, R. (2000). Cortical processing of change detection: dissociation between natural vowels and two-frequency complex tones. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10590–10594. doi: 10.1073/pnas.180317297

Vroomen, J., and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cogn. Neurosci.* 22, 1583–1596. doi: 10.1162/jocn.2009.21308

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642. doi: 10.1017/S0048577299981908

Xu, W., Kolozsvari, O. B., Oostenveld, R., Leppänen, P. H. T., and Hamalainen, J. A. (2019). Audiovisual processing of chinese characters elicits suppression and congruency effects in MEG. *Front. Hum. Neurosci.* 13:18. doi: 10.3389/fnhum.2019.00018

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage* 46, 226–240. doi: 10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., and Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage* 26, 703–720. doi: 10.1016/j.neuroimage.2005.02.040

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 25 July 2019
doi: 10.3389/fnhum.2019.00256

Action Semantic Deficits and Impaired Motor Skills in Autistic Adults Without Intellectual Impairment

Josephina Hillus[1], **Rachel Moseley**[2], **Stefan Roepke**[1] **and Bettina Mohr**[1]*

[1]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin, Germany,
[2]Department of Psychology, Bournemouth University, Poole, United Kingdom

Several studies indicate the functional importance of the motor cortex for higher cognition, language and semantic processing, and place the neural substrate of these processes in sensorimotor action-perception circuits linking motor, sensory and perisylvian language regions. Interestingly, in individuals with autism spectrum disorder (ASD), semantic processing of action and emotion words seems to be impaired and is associated with hypoactivity of the motor cortex during semantic processing. In this study, the relationship between semantic processing, fine motor skills and clinical symptoms was investigated in 19 individuals with ASD and 22 typically-developing matched controls. Participants completed two semantic decision tasks involving words from different semantic categories, a test of alexithymia (the Toronto Alexithymia Scale), and a test of fine motor skills (the Purdue Pegboard Test). A significant Group × Word Category interaction in accuracy ($p < 0.05$) demonstrated impaired semantic processing for action words, but not object words in the autistic group. There was no significant group difference when processing abstract emotional words or abstract neutral words. Moreover, our study revealed deficits in fine motor skills as well as evidence for alexithymia in the ASD group, but not in neurotypical controls. However, these motor deficits did not correlate significantly with impairments in action-semantic processing. We interpret the data in terms of an underlying dysfunction of the action-perception system in ASD and its specific impact on semantic language processing.

Keywords: autism, semantic processing, language, motor, action words

OPEN ACCESS

Edited by:
Xiaolin Zhou,
Peking University, China

Reviewed by:
Rutvik H. Desai,
University of South Carolina,
United States
Lihui Wang,
Otto von Guericke University
Magdeburg, Germany

****Correspondence:***
Bettina Mohr
bettina.mohr@charite.de

Received: *10 April 2019*
Accepted: *09 July 2019*
Published: *25 July 2019*

Citation:
Hillus J, Moseley R, Roepke S and Mohr B (2019) Action Semantic Deficits and Impaired Motor Skills in Autistic Adults Without Intellectual Impairment.
Front. Hum. Neurosci. 13:256.
doi: 10.3389/fnhum.2019.00256

INTRODUCTION

Neuroscientific research on "embodied cognition" postulates that higher cognitive processes, such as language, thought and reasoning, are functionally (and possibly structurally) interwoven with lower-level sensory and motor functions (Gallese and Lakoff, 2005; Barsalou, 2010). To this end, recent empirical evidence from behavioral and neuroimaging studies demonstrate that the motor cortex serves an important function for language processing, particularly during semantic processing (Pulvermüller, 1999; Pulvermüller et al., 2005; Moseley et al., 2013). More specifically, semantic processing of words associated with actions and motor movements activate the motor cortex somatotopically (Hauk et al., 2004; Pulvermüller and Fadiga, 2010; Moseley et al., 2012), which may be explained on the basis of the formation and activation of sensorimotor action-perception circuits comprising neurons in the motor cortex, in sensory cortices and in perisylvian language areas (Pulvermüller and Fadiga, 2010; Pulvermüller, 2012; Pulvermüller et al., 2014). Interestingly, recent data reveal a specific weakness in the processing of action-related words in clinical populations who have motor impairments (Boulenger et al., 2009; Bak and Chandran, 2012; Fernandino et al., 2013a,b; Cardona et al., 2014; Kemmerer, 2014; Desai et al., 2015). Specific impairments in action-semantic processing have also been reported in individuals with autism spectrum disorder (ASD), a neurodevelopmental syndrome characterized by problems with social interaction, communication and language, and, importantly, by dysfunction in motor behavior [American Psychiatric Association (APA), (2000)]. The motor deficits seen in ASD, ranging from differences in gait, fine motor skills, posture and coordination, are pervasive across the spectrum, occur in individuals with and without intellectual impairment, and are among the earliest symptoms to appear (Leary and Hill, 1996; Jansiewicz et al., 2006; Dziuk et al., 2007; Ming et al., 2007; Moseley and Pulvermüller, 2018). Unsurprisingly, abnormalities in structural and functional connectivity have been reported within and between primary motor cortex and other cortical regions in ASD (Mostofsky et al., 2007, 2009; McCleery et al., 2013; Floris et al., 2016; Thompson et al., 2017), as have differences in gray matter volume (Duffield et al., 2013; Mahajan et al., 2016), thus suggesting that the action-semantic deficit in this group is comparable to that seen in other populations with disease or damage to the motor system.

In the past, cognitive theories of ASD have centered around the archetypal "autistic triad" of deficits in social interaction, social communication and social imagination (Wing and Gould, 1979); as such, obvious motor impairments have been traditionally regarded as secondary and consequently neglected in research. To date, few studies on autism have focused on highlighting the functional relationship between motor symptoms and difficulties in higher-order cognitive functions, which include action-related cognition (e.g., imitation and gesturing). The functional link between an observed action and its corresponding motor program may be required to perform a self-generated movement and has been attributed to the *mirror neuron system* (MNS) which is posited to exist across primary and premotor cortex, somatosensory cortex, and parietal cortex. Responsive to both action perception and action execution, mirror neurons appear to be a quintessential type of multimodal "information-mixing" neuron, and a crucial element in binding motor areas to sensory and perisylvian language areas in action-perception circuits (Moseley and Pulvermüller, 2018). A number of studies consequently suggest that the MNS may be relevant in action perception, imitation, prediction of goals and intentions, as well as in social cognition and language (Iacoboni, 2009; Rizzolatti and Sinigaglia, 2010).

Previous studies have demonstrated functional impairments and neuronal hypoactivity of the MNS in autism (Nishitani et al., 2004; Oberman et al., 2005; Iacoboni and Dapretto, 2006; Bernier et al., 2007; Cattaneo et al., 2007; Honaga et al., 2010; Rizzolatti and Fabbri-Destro, 2010; McCleery et al., 2013; Wadsworth et al., 2017). These are consequently posited as the neuronal substrate of behavioral deficits in action-related cognition, which are interpreted as a consequence of dysfunctional action-perception mapping. This is manifest in impaired semantic processing for action but not object words in autistic individuals without intellectual disability, an impairment which correlated with reduced activation in cortical motor regions during action-word processing (Moseley et al., 2014, 2015; Moseley and Pulvermüller, 2018). Moreover, further studies in this clinical group revealed hypoactivation in motor as well as in limbic areas during processing of abstract emotional words (Moseley et al., 2012, 2015), which other studies have shown to be a notable challenge for autistic people. These findings have been interpreted on the basis that both of these semantic categories (action and emotion words) typically involve the activation of premotor and motor action-perception networks during learning and require this activity for efficient, optimal comprehension. This is consistent with the recent suggestion that hypoactivity of the motor cortex could also be one of the reasons for deficits in the socio-communicative and emotional-affective domain in ASD (Mody et al., 2017). Functional impairments between the motor cortex and perisylvian language regions may thus be related to social-communicative and emotional-affective deficits in individuals with ASD, as the development of semantic concepts would be mandatory for verbally expressing and understanding emotions in oneself and others.

A different theoretical approach explains reduced comprehension of emotional stimuli in ASD in terms of alexithymia, a difficulty in expressing and identifying one's own emotional states or feelings (Silani et al., 2008; Milosavljevic et al., 2016; Gaigg et al., 2018). However, a point of convergence might be that alexithymia itself may be (partially) caused by dysfunctional semantic processing of emotion words, which might, in turn, be linked to impaired action-perception circuits involving motor and limbic regions. Emotions clearly influence the style in which an action is performed, and thus predictably, the same multimodal mirror neurons of frontal-motor and parietal cortex are sensitive to different emotional states underpinning the same observed action (Di Cesare et al., 2015). This suggests the importance of the motor system in perceiving emotional states.

Previous studies demonstrated atypical brain activity in motor systems whilst autistic people read action and emotion words (Moseley et al., 2014, 2015), which also seems to be linked to a behavioral slowness in processing action words (Moseley et al., 2013). The next piece of this puzzle, however, remains missing: the link between language impairment for action and emotion words and *movement* impairment. To clarify this functional link, our study aimed to investigate the relationship between semantic processing of action and emotion words, fine and gross motor skills, and clinical symptoms in individuals with ASD and in typically-developed (TD) controls. In line with previous research with autistic participants, we predicted a specific processing deficit for action and emotion words but no groups differences for other word categories. We hypothesized that deficits in motor skills in individuals with ASD would be associated with clinical symptoms and impairments in processing these specific word categories.

MATERIALS AND METHODS

Participants

Nineteen autistic adults without intellectual disability (seven women) and 23 TD controls (nine women) were recruited for the study. One control participant had to be excluded from the final analysis due to poor task performance in the semantic decision task; therefore, the final data set comprised 19 ASD and 22 TD participants. All participants had normal or corrected-to-normal vision. In the control group, none of the participants had a history of psychiatric illness. Three participants in the ASD group took antidepressants.

The groups were matched for age, education, non-verbal IQ (measured by the LPS-3, Horn, 1983), and handedness (measured by the Edinburgh Handedness Inventory, Oldfield, 1971). Except for two participants in the ASD group, all participants were right-handed with a matched laterality-quotient (LQ). All participants were monolingual, native speakers of German. More information on both groups can be found in **Table 1**.

All ASD participants were diagnosed and recruited from the Autism Outpatient Clinic at the Charité University Medical School, Benjamin Franklin Campus, Berlin, Germany. Autism-specific diagnostic instruments were used for diagnosis, including the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2002) and a semi-structured clinical interview based on ASD criteria in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition [DSM-IV; American Psychiatric Association (APA), (2000)]. If a parent was available—which was the case in 66% of all ASD patients—the Autism Diagnostic Interview-Revised (Lord et al., 1994) was conducted. Final diagnoses were established by expert consensus taking into account clinical interviews and scale assessments. A patient was diagnosed with ASD when scores on both the ADOS and the ADI-R exceeded the cut-off for autism spectrum or autism and all required DSM-IV criteria of the clinical interview were fulfilled. For the 33% of patients whose parents were not available for the ADI-R interview, an ASD diagnosis was given when all required criteria of the ADOS and the clinical interview were met and the patient provided sufficient examples that the autistic symptoms already existed in childhood.

TABLE 1 | Means and standard deviations (SD, in brackets) of demographic and clinical variables used to match the autism spectrum disorder (ASD) and TD groups.

	ASD group *N* **= 19**	**TD control group** *N* **= 22**	**Statistical group difference**
Age (years)	39.00 (11.20)	36.59 (7.55)	n.s. ($p = 0.4$)
Education (years)	12.00 (1.52)	12.73 (0.88)	n.s. ($p = 0.06$)
IQ (LPS-3)	117.76 (9.75)	112.96 (8.72)	n.s. ($p = 0.1$)
Laterality Quotient (LQ)	79.79 (16.09)	88.18 (15.31)	n.s. ($p = 0.09$)
Autism-Spectrum Quotient (AQ)	39.05 (6.62)	11.59 (4.02)	$p < 0.001$

Between-group differences were calculated by independent t-tests (p-values are in brackets; n.s. indicates non-significant result). Groups did not differ on any variable except on the AQ.

The mean score of the ASD group on the Autism-Spectrum Quotient (AQ: Baron-Cohen et al., 2001) was 39.1 (SD: 6.6) compared to a mean score of 11.59 (SD: 4.020) in the control group: as expected, a significantly higher average score ($t_{(39)} = 16.302$, $p < 0.001$). All but one participant in the ASD group scored above 26, which is considered as the general cut-off point for diagnosable autism (Woodbury-Smith et al., 2005).

Neuropsychological and Clinical Assessment

Leistungsprüfsystem-Test, Subtest 3

The *Leistungsprüfsystem-Test, Subtest 3* (Horn, 1983) was carried out with all participants to assess non-verbal IQ. Handedness was measured by the *Laterality Quotient*, assessed by the *Edinburgh Handedness Inventory* (Oldfield, 1971).

Purdue Pegboard Test

The Purdue Pegboard Test was used in both groups to assess manual dexterity, manual coordination and fingertip skills (Tiffin and Asher, 1948). The test consists of a board with two parallel rows of 25 holes running vertically. Participants were asked to use their right hand to put as many of the cylindrical metal pegs as possible in the right-sided row within 30 s; the same procedure was then followed for the left hand with the left-sided row. In a third condition which combined the two previous trials, participants had to simultaneously place the pegs within the right- and left-sided rows with their right and left hands respectively. In a fourth condition, as many "assemblies" as possible, consisting of different objects, had to be built within 60 s.

Trailmaking Test (Parts A and B)

The Trailmaking Test (TMT; Parts A and B) is a neuropsychological test to measure attention, processing speed and executive functions (Tombaugh, 2004). This test was performed with the ASD group only in order to assess psychomotor speed and attention (Part A) as well as executive function (Part B).

Clinical Questionnaires

All participants filled out the Autism-Spectrum Quotient (AQ) and the Toronto Alexithymia Scale 26 (TAS-26; Taylor et al., 1985). The AQ measures the degree of autistic traits whereby higher scores indicate a higher degree of autistic traits (Baron-Cohen et al., 2001). This most popular dimensional measure of autistic traits has been extensively used and validated both in the general population and those with diagnosed autism (Hurst et al., 2007; Hoekstra et al., 2008; Ruzich et al., 2015, 2016; Stevenson and Hart, 2017), where it boasts sound psychometric properties.

Alexithymia is popularly understood as a dimensional construct (Keefer et al., 2019) which is most commonly measured with the TAS-26. This scale comprises three subscales assessing the difficulties describing emotions (scale 1), difficulties identifying one's own emotions (scale 2), and the tendency to think in an externally-oriented way (scale 3).

Furthermore, all ASD participants filled out the Empathy Quotient (EQ; Baron-Cohen et al., 2014) and the Systemizing Quotient-R (SQ-R; Baron-Cohen et al., 2003; Wheelwright et al., 2006). The EQ measures the capacity for empathy, whereby a lower score indicates reduced empathy. The SQ-R measures the capacity for recognizing patterns and the tendency to "systemize," to see the world in terms of logical rules and systems and to try to impose these in life, whereby higher scores reflect greater tendency to systemizing. Developed by the same group as the AQ, EQ scores tend to be lower and SQ-R scores higher in autistic individuals, and both short forms of the original tests showed good psychometric properties (Wheelwright et al., 2006).

In an additional, self-designed questionnaire, the MOSES-Test ("Motor Skills in Everyday Situations"), participants had to self-assess their motor skills in everyday situations on a four-point Likert scale employing 12 statements such as "I can easily catch or throw a ball," or "I have no difficulties riding a bike." Possible scores ranged from 0 ("I completely agree") to 3 ("I completely disagree"). If the statements concerned difficulties ("I have difficulties in climbing stairs"), then scores ranged from 3 ("I completely agree") to 0 ("I completely disagree"). With an upper limit of 36, higher scores on this questionnaire suggest more difficulties in gross motor skills. The MOSES-Test can be found in the **Supplementary Materials**.

Semantic Decision Tasks

Stimuli

In the first semantic decision task (SDT1; see details below), 90 action-related words {30 face-related [e.g., "BEISSEN" ("TO BITE")], 30 hand-related [e.g., "MALEN" ("TO PAINT")], 30 foot-related [e.g., "LAUFEN" ("TO WALK")]} and 90 object-related words {30 animal words [e.g., "MAUS" ("MOUSE")], 30 tool words [e.g., "HAMMER" ("HAMMER")], 30 food words [e.g., "KUCHEN" ("CAKE")]} were included.

In the second semantic decision task (SDT2; details below), we included 30 abstract emotional words [e.g., "FREUDE" ("JOY")] and 30 abstract neutral words [e.g., "PLANEN" ("TO PLAN")]. Abstract emotional words consisted of verbs and nouns associated with emotions, and the abstract neutral word category included verbs and nouns referring to emotionally neutral concepts or cognitions. Words were selected and matched as carefully as possible based on psycholinguistic properties such as word length and word frequency according to the CELEX database (Baayen et al., 1993).

Before conducting this experiment, a semantic rating study was carried out with 10 typically-developing participants who did not take part in the main experiment. This pre-experiment rating study was conducted to differentiate the selected word categories with respect to their semantic properties (see also Hauk et al., 2004; Moseley et al., 2015). Study participants rated all words with regards to semantic features such as concreteness, arousal, valence, emotion-relatedness and action-relatedness. Psycholinguistic variables and semantic ratings for the four major stimulus categories (action-, object-, abstract emotional-, abstract internal words) used in SDT 1 and 2 are displayed in the **Supplementary Materials**.

Procedure

All participants performed two separate and independent semantic decision tasks (SDT1 and SDT2) using E-prime software (Psychology Software Tools, Inc., Sharpsburg, PA, USA, RRID:SCR_009567). The first SDT1 was carried out employing action- and object-related words; the second SDT2 task used abstract emotional and abstract neutral words. Each semantic decision task lasted 10 min, with a break given in between.

Participants were seated approximately 60 cm distance from the computer screen while words appeared on a white background in uppercase, black bold print. All participants were asked to decide as fast and accurately as possible if the presented words were related to human actions or to objects (in SDT1) or, in the second task (SDT2), whether the words were related to emotional or non-emotional abstract concepts. Participants indicated their semantic judgments by pressing one of two keys on a computer keyboard with the index and middle fingers of their right hand. The assignment of keys was counterbalanced between participants. After a fixation cross was shown at central location for 250 ms, words were presented tachistoscopically for 150 ms in a pseudorandomized order. Participants were shown the same words with each word being only shown once to each participant. After the offset of the word, a blank screen was shown until the participant made a decision, or until 2,500 ms had passed without a response, at which point the screen returned to the fixation cross. The stimulus onset asynchrony (SOA) was 2,500 ms. Instead of using their right hand, the two left-handed participants used the index and middle finger of their left hand to perform the SDTs.

Data Analysis

All data was analyzed using SPSS version 24.0 (RRID:SCR_002865). Independent *t*-tests were used to compare means of demographic variables, neuropsychological tests and clinical questionnaires.

For each participant, we derived mean reaction times and accuracy scores for each word category (action words and object words from SDT1, emotional and non-emotional abstract words from SDT2): this was done by averaging reaction times across all individual words in that category. Each word within a category received either a score of 1 (reflecting correct categorization) or 0 (reflecting that the participant had incorrectly categorized the word or failed to respond). For each participant, the means across these accuracy scores were then transformed into a percentage accuracy for each word category. As such, a mean accuracy score and a mean reaction time score for the action, object, abstract emotional and abstract non-emotional word categories were entered into SPSS for each participant.

To compare reaction times and accuracies of both groups for statistically significant differences, we performed four 2 × 2 mixed design repeated measures analysis of variances (ANOVAs). In each ANOVA, the between factor "Group" (two levels: ASD vs. control) and the within factor "Word Category" [two levels: action words vs. object words (SDT1), emotional vs. non-emotional abstract words (SDT2)] were included.

As concepts, tools and the words denoting them are known to evoke activity in motor regions which are associated with their action affordances, i.e., the actions associated with their use (Chao and Martin, 2000; Carota et al., 2012). As such, these "object-related" tool words tend to be semantically related not only to visual objects but also to specific actions (for instance, a fork to eating). In order to control for this potential "action-relatedness" of the tool word category, we conducted another ANOVA in which tool words were excluded from the analysis. *Post hoc* planned comparisons were conducted with subsequent Bonferroni corrections.

A Pearson correlation was computed for each group separately to assess the relationship between accuracy and latency for each word category in the semantic decision tasks and other variables (AQ, TAS-26, EQ, SQ-R and MOSES-Test). No outlier removal procedure was applied as none of the individual data sets exceeded the mean group values by more than two standard deviations.

RESULTS

Neuropsychological and Clinical Assessment

Purdue Pegboard

T-tests revealed significant differences between the two groups in the first three conditions of the Purdue Pegboard Test (PPB), but not in the fourth "assembly" condition. In comparison to the control group, the ASD group placed significantly fewer pegs with their right hands, left hands and with both hands simultaneously, thus demonstrating impaired fine motor skills (see **Table 2**).

Trailmaking Test A and B

We conducted the TMT A and B only for the ASD group and found a mean of 22.05 s (SD: 7.50) in the TMT A and a mean of 49.58 s (SD: 17.58) in the TMT B, indicating unimpaired performance in the range of norms from healthy participants as stated in the test.

Clinical Questionnaires

Toronto-Alexithymia-Scale-26

T-tests showed a significant difference between the ASD group and the TD group in overall TAS-26 scores (see **Table 3**) and in all three sub-scales.

EQ and SQ-R

The Empathy Questionnaire (EQ) and the Systemizing Questionnaire Revised (SQ-R) were only filled out by the ASD group. The mean score on the SQ-R was 79.21 (SD: 22.837). The mean score of the EQ was 13.89 (SD: 5.597) which is comparable (even slightly lower) than the empathy scores seen in the autistic sample of the original and certainly under the recommended cut-off score of 30, which allowed the authors to distinguish 81% of their autistic sample (Baron-Cohen and Wheelwright, 2004).

MOSES-Test

A t-test revealed a significant difference in the overall MOSES-score between the ASD group and the control group ($p < 0.001$). The ASD group scored significantly higher with a mean score of 14.53 (SD: 6.851) compared to a mean score of 4.50 (SD: 2.956) in the control group, indicating more motor difficulties in everyday life situations.

Semantic Decision Tasks

SDT1: Action Words vs. Object Words

A mixed-design repeated measures ANOVA revealed a significant *Group* × *Word Category* interaction for accuracy ($F_{(1,39)} = 4.01$, $p < 0.05$, $\eta^2_{\mathrm{p}} = 0.093$; see **Figure 1**). *Post hoc* analyses using pairwise comparisons (Bonferroni-corrected) showed that participants in the ASD group made significantly more errors when presented with action words than they did to object words ($p < 0.05$). This interaction did not show significance in the latency analysis ($F_{(1,39)} = 0.0001$, $p = 0.985$, $\eta^2_{\mathrm{p}} = 0.0003$). There was no significant main effect of *Group* in accuracy $F_{(1,39)} = 2.42$, $p = 0.128$, $\eta^2_{\mathrm{p}} = 0.06$) or latency $F_{(1,39)} = 0.88$, $p = 0.355$, $\eta^2_{\mathrm{p}} = 0.02$), suggesting that where differences did appear, they were associated with particular word categories rather than generally poorer or slower processing. However, a significant main effect of *Word Category* in the latency analysis ($F_{(1,39)} = 27.15$, $p < 0.001$, $\eta^2_{\mathrm{p}} = 0.41$) suggested that *all* participants were slower to process action words; there was a non-significant tendency for them to be less accurate for action words, too ($F_{(1,39)} = 2.87$, $p = 0.098$, $\eta^2_{\mathrm{p}} = 0.07$). Means for accuracies and latencies are presented in **Table 4**.

Furthermore, sub-categories of object and action words were investigated in *post hoc* analyses applying Bonferroni-corrected pairwise comparisons. The analyses revealed that in the control group, there were significant differences between animal words and tool words ($p = 0.001$), between tool words and food words ($p = 0.002$), and between animal words and each effector-specific type of action word (face-related words: $p < 0.001$; hand-related words: $p < 0.001$; foot-related words: $p < 0.001$). In the ASD group, there were only significant differences between animal

TABLE 2 | Means, standard deviations (in brackets) and statistical group comparisons in the Purdue Pegboard (PPB) Test.

	ASD group $N = 19$	Control group $N = 22$	Statistical testing (t)
PPB right	14.16 (1.53)	15.77 (1.51)	$p < 0.01$
PPB left	13.42 (2.38)	14.82 (1.43)	$p < 0.05$
PPB both	11.47 (1.57)	12.41 (1.26)	$p < 0.05$
PPB Assembly	34.74 (7.43)	36.41 (6.68)	n.s. ($p = 0.45$)

Statistically significant effects are indicated by p-values; n.s. indicates non-significant difference.

TABLE 3 | Means, standard deviations (in brackets) and statistical group comparisons in the TAS-26 questionnaire.

	ASD group $N = 19$	Control group $N = 22$	Statistical testing (t)
TAS-26	49.00 (10.29)	38.09 (5.97)	$p < 0.001$
TAS-26 (Scale 1)	18.53 (6.51)	12.09 (2.94)	$p < 0.001$
TAS-26 (Scale 2)	17.79 (4.34)	11.64 (3.65)	$p < 0.001$
TAS-26 (Scale 3)	12.68 (2.81)	14.36 (2.57)	$p < 0.05$

Statistically significant effects are indicated by p-values.

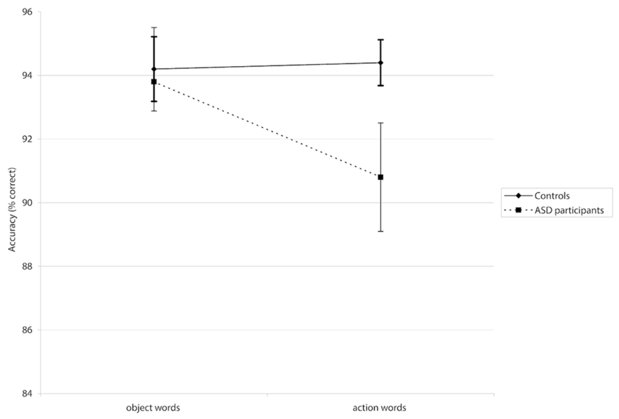


FIGURE 1 | The *Word Category* × *Group* interaction is displayed for the semantic decision task in which action words and object words were presented. The error bars show the standard error of the mean. Autism spectrum disorder (ASD) participants showed significantly lower accuracy specifically for action words than controls. There was no sign. difference between both groups when processing object words.

TABLE 4 | Means and standard deviations (in brackets) for latencies and accuracies.

	ASD group	Control group
I Action words—Object words		
Reaction time (ms) Action words	630.09 (188)	590.26 (121)
Reaction time (ms) Object words	573.58 (134)	533.34 (115)
Accuracy (%) Action words	90.8 (7.4)	94.4 (3.1)
Accuracy (%) Object words	93.8 (4.0)	94.2 (4.4)
II Abstract emotional words—Abstract neutral words		
Reaction time (ms) Abstract emotional words	816.90 (379)	618.11 (136)
Reaction time (ms) Abstract neutral words	885.61 (374)	774.62 (208)
Accuracy (%) Abstract emotional words	91.90 (9.4)	95.80 (4.4)
Accuracy (%) Abstract neutral words	81.70 (14.5)	90.70 (8.1)

words and tool words (p = 0.005) and between animal words and foot-related action words (p = 0.011), but not between animal words and the other effector-specific action words (hand-related or face-related words), or between tool words and food words.

SDT2: Abstract Emotional vs. Abstract Neutral Words

The ANOVA revealed a main effect of *Word Category* both in accuracy ($F_{(1,39)}$ = 14.38, p < 0.001, η_p^2 = 0.26) and latency ($F_{(1,39)}$ = 16.69, p < 0.001, η_p^2 = 0.30): in both cases, all participants were faster and more accurate for abstract emotional than abstract neutral words. Furthermore, there was a significant main effect of *Group* in the accuracy analysis ($F_{(1,39)}$ = 8.25, p = 0.007, η_p^2 = 0.17) with significantly fewer correct responses for all words, regardless of word category, in the ASD group (see **Table 4**). No significant main effect of *Group* was found in the latency analysis ($F_{(1,39)}$ = 3.28, p = 0.078, η_p^2 = 0.08). Moreover, there was no significant *Group* × *Word Category* interaction for accuracy ($F_{(1,39)}$ = 1.66, p = 0.205, η_p^2 = 0.04) or latency ($F_{(1,39)}$ = 2.54, p = 0.119, η_p^2 = 0.06), suggesting no particular category-specific deficit specific to either group.

Correlations Between Clinical Data and Semantic Decisions

Pearson correlations were performed between neuropsychological tests, clinical scales, and latency and accuracy data from the semantic decision tasks. The results showed a positive correlation in the ASD group between AQ scores and the overall TAS-26 score (r = 0.674, p = 0.002). Furthermore, in the ASD group, there was a positive correlation between AQ scores and the MOSES-Test (r = 0.766, p < 0.001). Regarding the EQ, a negative correlation between AQ and the EQ scores in the autistic group (r = −0.499, p = 0.03) corroborated previous research, where higher scores on the AQ were associated with lower scores on the EQ. However, there was no significant correlation between any of these tests and the accuracy or latency of semantic judgments for any particular word category.

DISCUSSION

This study aimed to elucidate the relationship between semantic processing, motor skills and clinical variables in autistic individuals and IQ-matched neurotypical controls. In line with previous findings of action word deficits (Moseley et al., 2013), a significant Group × Word Category interaction was found for accurate data and revealed that autistic participants were significantly less accurate than typically-developing controls when processing words associated with actions. Importantly and in contrast, the ASD group performed as accurately as controls when making semantic decisions about object-related words. This category-specific deficit in action-semantic processing, seen here in another motor-impaired group alongside those noted previously (Boulenger et al., 2009; Bak and Chandran, 2012; Fernandino et al., 2013a,b; Cardona et al., 2014; Kemmerer, 2014; Desai et al., 2015), might be interpreted in terms of an underlying dysfunction of the neuronal action-perception links (Rizzolatti and Fabbri-Destro, 2010; Moseley et al., 2013) suggested to underlie semantic processing (Pulvermüller and Fadiga, 2010; Moseley and Pulvermüller, 2018). Abnormalities in the circuits connecting motor regions to perisylvian language cortices would result in difficulties recognizing or understanding those words which especially draw on these links for the motor programs supporting conceptual knowledge: namely, in the first instance, action words (for a comprehensive review, see Moseley and Pulvermüller, 2018). It is important to note the specificity of this action-semantic processing deficit in the present and the previous study (Moseley et al., 2013), which speaks against the assumption of a more generic semantic language impairment in ASD, which might have been reflected by main effects of

Group in SDT1 (see below for discussion of SDT2). Previous studies suggest that the weakness that some clinical groups show in processing action-related stimuli is related to the differing semantic content of action-words and object-related words, rather than their differing grammatical roles (Pulvermüller and Fadiga, 2010; Moseley and Pulvermüller, 2018).

In support of the notion of an underlying action-motor problem in ASD, we found evidence for impaired motor skills in the ASD group compared to controls: in the Purdue Pegboard Test, the ASD group showed reduced hand motor skills when placing pegs in a board with the left hand, the right hand, and with both hands simultaneously. Interestingly, when a complex assembly of different objects with both hands was required, control participants and individuals with ASD performed equally well. Besides fine motor skills, the assembly task tests for bimanual coordination and executive function: our results may suggest that our autistic sample were able to compensate for deficits in unimanual fine motor skills by good performance on bimanual coordination. Although executive dysfunction in autism is assumed to be evident in everyday functioning, it is difficult to capture experimentally in tests with low ecological validity (Kenworthy et al., 2008; Wallace et al., 2016) and poor sensitivity (Demetriou et al., 2018). "Executive function" is a term which encapsulates many higher-level processes, and autistic people tend to show a somewhat inconsistent performance of executive difficulties and executive sparing, which is affected by sample differences in age, gender, IQ (where, notably, our study included only individuals with IQ in the normal range), by common comorbidities such as depression, anxiety and ADHD, and by task features such as complexity, whether tasks are open-ended or more structured (Demetriou et al., 2018) or even whether they measure cognitive performance vs. overt manifestations of difficulties (Albein-Urios et al., 2018)[1]. It is highly likely that the lack of executive impairment seen in our data belies significant difficulties in everyday life (Wallace et al., 2016).In this context, it seems not especially surprising that the autistic sample in our study did not appear impaired on the TMT Parts A and B, where they were compared with normative data from typically-developing participants in the same age range (Tombaugh, 2004). In contrast to previous studies (Hill and Bird, 2006), individuals with ASD in our study performed well on both parts of the TMT, though we were unable to perform a direct comparison to our own control group who did not complete the TMT. Interestingly and specifically relating to the TMT, a stronger performance has been seen in autistic girls and women than autistic boys and men (Bölte et al., 2011; Lehnhardt et al., 2016). This may furthermore explain a lack of group differences in our sample of men *and* women.

To our knowledge, this study is the first one to employ a semantic decision task with abstract emotional and abstract but emotionally neutral words. Based on previous data demonstrating cortical hypoactivation in the motor and limbic cortex in individuals with ASD when processing emotion words (Moseley et al., 2015) and data from patients with motor lesions (Dreyer et al., 2015), we expected to find evidence for impaired processing of abstract emotional words but not for emotionally neutral abstract words; these, like action words, would draw on motor systems for meaning (Moseley et al., 2012) and thus be especially impaired in our participants with movement impairments. Our data did not confirm this prediction but revealed that the ASD group, in general, showed less accurate and slower performance than typically-developing controls, irrespective of these two-word categories. One possible explanation of this finding could be due to the fact that the SDT2 task (abstract emotional words vs. abstract neutral words) was more difficult than the SDT1 task (action vs. object words). This might have led to a lower and more heterogeneous performance in the SDT2 task in both groups, reducing statistical power and thus working against the emergence of a statistically significant Group × Word category interaction.

Correlation analyses calculated between neuropsychological and clinical tests and accuracy and reaction time for semantic decisions did not reveal any statistically significant relationships, including (most notably for this study) a lack of relationship between movement impairments (in both the Purdue Test AND the MOSES-Test) and reaction times and accuracy for those word categories hypothesized to depend most on motor systems: action words and abstract emotional words. As such, our original hypothesis, that autistic deficits in motor skills would be functionally associated with impairments in action-semantic processing, was not statistically supported by the data. This is unexpected given the relationship between motor hypoactivity and impaired action word processing seen previously (Moseley et al., 2013). This previous study in autism, as well as reports from other patient groups with diseases or lesions of the motor system (Boulenger et al., 2009; Bak and Chandran, 2012; Cardona et al., 2014; Kemmerer, 2014), suggest the functional importance of the motor system for optimal action word processing; the studies above also indicate a functional role for motor systems for abstract emotional words (Moseley et al., 2012, 2015; Dreyer et al., 2015) though this proposition has not yet accrued the same degree of empirical support. For action words, at least, simulation studies and studies of novel action word learning have been able to demonstrate the involvement and importance of motor systems in acquiring an action vocabulary. The fact that action and emotion word processing deficits were not related to motor dysfunction appears to speak against this interpretation. However, an interesting possibility is whether the deficits in hand dexterity shown here by the Pegboard Test may have been so specific that they did not correlate with errors to action words which ranged in effector-specificity, as the overall action word category included not only hand-related action words that might correspond with the motor programs employed by the Purdue

---

[1] Indeed, with reference to heterogeneity in task performance, it is important to note that although our autistic sample showed motor deficits in the majority of conditions in the Purdue Pegboard Test, other findings range from an absence of any impairments (Lai et al., 2012), impairments across the board (Barbeau et al., 2015), or inconsistent profiles contradictory to our sample (for instance, poorer performance in the assembly and right-handed condition, but not in the left-handed and simultaneous bimanual condition (Thompson et al., 2017). Again, it should be noted that motor skills are likewise affected by participant characteristics such as autistic symptom severity, IQ, language development and age, and the influence of sex is so far unknown (Moseley and Pulvermüller, 2018).

Pegboard Test, but also those denoting motor programs of the feet and face. The same point could be made regarding emotion words, which foremost tend to be related to actions of the face (Moseley et al., 2012). A more thorough investigation might, as such, include a wider battery of motor tests and a larger sample size with greater power. It is also notable that autistic individuals may, to some extent, be able to compensate for impaired motor systems by recruiting other areas for semantic word processing (Moseley and Pulvermüller, 2018). This may be another reason for the lack of an association, and ultimately, studies would benefit from marrying multiple methodologies: imaging during language testing, *and* motor skills testing.

A notable limitation of our study is the fact that semantic differences between action and object words were confounded by uncontrolled differences in grammatical class: action words were all verbs, while object words were nouns which could have confounded our data. As such, it could be argued that autistic participants had a general deficit across the grammatical category of verbs. Though this study cannot speak to this possibility, our previous investigation in autistic participants found a double dissociation *within* the grammatical category of verbs between words with emotional content and those without (Moseley et al., 2015). Analysis of carefully orthogonalized word categories does indeed suggest that action and object words diverge along the semantic as opposed to grammatical line (Moseley and Pulvermüller, 2014), though dissociations between nouns and verbs as grammatical categories might appear as emergent properties of the more fundamental difference in action and object associations. The primacy of the semantic as opposed to grammatical dissociation has been supported by a number of studies (Barber et al., 2010; Vigliocco et al., 2011; Kemmerer et al., 2012; Fargier and Laganaro, 2015; Lobben and D'Ascenzo, 2015; Popp et al., 2016; Zhao et al., 2017; Vonk et al., 2019), though others reflect both semantic *and* grammatical divisions (Yudes et al., 2016; Yang et al., 2017). We would as such doubt that our findings reflect a general verb deficit in autism, but as debate surrounding the amodal vs. modal organization of language continues, we cannot speak conclusively on this matter.

Another point of note is that one of our subcategories of object words, tool words, is known to elicit activity in motor systems that has been associated with the action affordances of these objects (Chao and Martin, 2000; Carota et al., 2012). Including this more action-related subcategory within our superordinate object-word category might, therefore, have been problematic. In an attempt to exclude the possible contribution of action associations from tool words in our object word category, we ran a secondary analysis excluding tool words, which did not lead to a different pattern of results. As such, the autistic impairment seen for action words was impervious to the presence of tool words in the object word category, but along with tighter control over the grammatical confound of action/verbs and object/nouns, future studies may wish to exclude tool words within superordinate object word categories.

Whilst none of the motor or clinical tests correlated with the semantic language tasks, several other relationships of interest were observed which corresponded with previous research in autism. First, a significant correlation between the severity of autistic symptoms (as measured by the AQ) and the severity of alexithymia (as measured by the TAS-26) was obtained in our autistic participants. This finding suggests that a higher number of autistic traits is associated with greater alexithymia, and is in line with other research that has shown high comorbidity between ASD and alexithymia (Lombardo et al., 2007; Milosavljevic et al., 2016; Kinnaird et al., 2019). Our ASD participants had significantly higher overall scores on all scales of the TAS-26 in comparison to TD controls. Scale 1 of the TAS-26 measures difficulties in identifying feelings, scale 2 measures difficulties in describing (communicating) feelings, and scale 3 measures externally-orientated thinking.

A high degree of consistency was seen between our findings and previous literature on the AQ, the EQ, and the SQ-R: namely, that autistic participants had lower scores on the EQ and that empathy scores decreased as autistic traits increased (as in Baron-Cohen and Wheelwright, 2004; Wheelwright et al., 2006); and that as in previous studies, autistic individuals tend to score highly in systemizing (Baron-Cohen et al., 2003; Wheelwright et al., 2006). This pattern, overall, confirms the empathizing-systemizing account of autism (Baron-Cohen, 2009), and is consistent with that seen in very large samples (Baron-Cohen et al., 2014).

Our self-developed MOSES questionnaire evaluates problems in gross motor skills in daily life (e.g., catching a ball, riding a bicycle, descending stairs, standing on one leg). The ASD group scored significantly higher than controls on this self-report questionnaire, indicating gross motor deficits that corroborate the fine deficits seen in the Purdue Pegboard Test. Furthermore, there was a strong positive correlation between overall AQ scores and the MOSES questionnaire which implies that the degree of autistic traits may correspond to the severity of motor deficits in everyday life situations. Many studies have shown deficits in gross motor skills in individuals with ASD (Leary and Hill, 1996; Jansiewicz et al., 2006; Dziuk et al., 2007; Ming et al., 2007), and many studies have likewise shown a relationship between increased severity of autistic symptomatology and greater motor dysfunction (Papadopoulos et al., 2012; MacDonald et al., 2013, 2014; Travers et al., 2013, 2015; Stevenson et al., 2017; Uljarević et al., 2017; for review, see Moseley and Pulvermüller, 2018). Notably, the MOSES test in our study assessed how participants subjectively *perceived* their own gross motor skills. It is interesting that ASD participants' perception of their own deficits in gross motor function is consistent with the poorer scores in objective assessments of gross motor skills described in previous studies, and that as in previous studies, a relationship exists between motor deficits and autistic symptomatology, even when the former is self-reported.

Finally, this study possesses limited generalizability within the autism spectrum, due to the fact that only autistic adults without intellectual disability were included. Hence, these findings cannot be generalized to minimally-verbal adults, those with intellectual disability, or to children with ASD. Moreover, although the sample size in the present study is similar compared to other behavioral studies on autism, the results require confirmation in future studies with a larger clinical group.

CONCLUSION

Our study corroborates previous findings that autistic individuals show specific difficulties in semantic processing of action words; there was no evidence for differential semantic processing deficits for any other word category. Furthermore, our findings revealed deficits in fine motor skills as well as in self-reported gross motor behavior in autistic adults without intellectual disability. The results might be interpreted on the basis of impaired functional (or structural) connections within the motor cortex that hinders the formation of action-perception circuits which may be crucial for storing semantic concepts. The lack of a significant correlation between motor skills in ASD and deficits for action (and indeed emotion words) did not support the notion of a direct functional-behavioral link between motor performance and semantic processing of these words, but the study leaves open several possible interpretations. Further investigation is thus needed to corroborate the hypothesized functional relationship between motor deficits and impairments in processing words which imply motor regions.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Charité Ethics Committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Charité Ethics Committee.

AUTHOR CONTRIBUTIONS

JH contributed to the study design, recruitment and testing of participants, data analysis and writing of the manuscript. BM contributed to the study design, recruitment of participants, data analysis and writing of the manuscript. RM contributed to the study design and writing of the manuscript. SR contributed to the recruitment and testing of participants and writing of the manuscript.

FUNDING

The study was supported by Charité Universitätsmedizin Berlin. We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité—Universitätsmedizin Berlin.

ACKNOWLEDGMENTS

We thank our participants for taking part in this study. We are grateful to Verena Büscher, Friedemann Pulvermüller, Felix Dreyer, Alessandra Mancini, David Hillus, and Svenja Köhne for their help at various stages of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00256/full#supplementary-material

REFERENCES

Albein-Urios, N., Youssef, G. J., Kirkovski, M., and Enticott, P. G. (2018). Autism spectrum traits linked with reduced performance on self-report behavioural measures of cognitive flexibility. *J. Autism Dev. Disord.* 48, 2506–2515. doi: 10.1007/s10803-018-3503-3

American Psychiatric Association (APA). (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR).* Arlington, VA: American Psychiatric Association.

Bak, T. H., and Chandran, S. (2012). What wires together dies together: verbs, actions and neurodegeneration in motor neuron disease. *Cortex* 48, 936–944. doi: 10.1016/j.cortex.2011.07.008

Barbeau, E. B., Meilleur, A.-A. S., Zeffiro, T. A., and Mottron, L. (2015). Comparing motor skills in autism spectrum individuals with and without speech delay. *Autism Res.* 8, 682–693. doi: 10.1002/aur.1483

Barber, H. A., Kousta, S.-T., Otten, L. J., and Vigliocco, G. (2010). Event-related potentials to event-related words: grammatical class and semantic attributes in the representation of knowledge. *Brain Res.* 1332, 65–74. doi: 10.1016/j.brainres.2010.03.014

Baron-Cohen, S. (2009). Autism: the empathizing-systemizing (E-S) theory. *Ann. N Y Acad. Sci.* 1156, 68–80. doi: 10.1111/j.1749-6632.2009.04467.x

Baron-Cohen, S., Cassidy, S., Auyeung, B., Allison, C., Achoukhi, M., Robertson, S., et al. (2014). Attenuation of typical sex differences in 800 adults with autism vs. 3,900 controls. *PLoS One* 9:e102251. doi: 10.1371/journal.pone.0102251

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* 358, 361–374. doi: 10.1098/rstb.2002.1206

Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. doi: 10.1023/b:jadd.0000022607.19833.00

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Top. Cogn. Sci.* 2, 716–724. doi: 10.1111/j.1756-8765.2010.01115.x

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (1993). *The CELEX Lexical Database [CD-ROM].* Philadelphia, PA: Linguistics Data Consortium, University of Pennsylvania.

Bernier, R., Dawson, G., Webb, S., and Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain Cogn.* 64, 228–237. doi: 10.1016/j.bandc.2007.03.004

Bölte, S., Duketis, E., Poustka, F., and Holtmann, M. (2011). Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. *Autism* 15, 497–511. doi: 10.1177/1362361310391116

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914. doi: 10.1093/cercor/bhn217

Cardona, J. F., Kargieman, L., Sinay, V., Gershanik, O., Gelormini, C., Amoruso, L., et al. (2014). How embodied is action language? Neurological evidence from motor diseases. *Cognition* 131, 311–322. doi: 10.1016/j.cognition.2014.02.001

Carota, F., Moseley, R., and Pulvermüller, F. (2012). Body-part-specific representations of semantic noun categories. *J. Cogn. Neurosci.* 24, 1492–1509. doi: 10.1162/jocn_a_00219

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., et al. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proc. Natl. Acad. Sci. U S A* 104, 17825–17830. doi: 10.1073/pnas.0706273104

Chao, L. L., and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–484. doi: 10.1006/nimg.2000.0635

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., et al. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Mol. Psychiatry* 23, 1198–1204. doi: 10.1038/mp.2017.75

Desai, R. H., Herter, T., Riccardi, N., Rorden, C., and Fridriksson, J. (2015). Concepts within reach: action performance predicts action language processing in stroke. *Neuropsychologia* 71, 217–224. doi: 10.1016/j.neuropsychologia.2015.04.006

Di Cesare, G., Di Dio, C., Marchi, M., and Rizzolatti, G. (2015). Expressing our internal states and understanding those of others. *Proc. Natl. Acad. Sci. U S A* 112, 10331–10335. doi: 10.1073/pnas.1512133112

Dreyer, F. R., Frey, D., Arana, S., von Saldern, S., Picht, T., Vajkoczy, P., et al. (2015). Is the motor system necessary for processing action and abstract emotion words? Evidence from focal brain lesions. *Front. Psychol.* 6:1661. doi: 10.3389/fpsyg.2015.01661

Duffield, T. C., Trontel, H. G., Bigler, E. D., Froehlich, A., Prigge, M. B., Travers, B., et al. (2013). Neuropsychological investigation of motor impairments in autism. *J. Clin. Exp. Neuropsychol.* 35, 867–881. doi: 10.1080/13803395.2013.827156

Dziuk, M. A., Larson, J. C. G., Apostu, A., Mahone, E. M., Denckla, M. B., and Mostofsky, S. H. (2007). Dyspraxia in autism: association with motor, social, and communicative deficits. *Dev. Med. Child Neurol.* 49, 734–739. doi: 10.1111/j.1469-8749.2007.00734.x

Fargier, R., and Laganaro, M. (2015). Neural dynamics of object noun, action verb and action noun production in picture naming. *Brain Lang.* 150, 129–142. doi: 10.1016/j.bandl.2015.09.004

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., et al. (2013a). Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain Lang.* 127, 65–74. doi: 10.1016/j.bandl.2012.07.008

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., et al. (2013b). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia* 51, 1510–1517. doi: 10.1016/j.neuropsychologia.2013.04.008

Floris, D. L., Barber, A. D., Nebel, M. B., Martinelli, M., Lai, M.-C., Crocetti, D., et al. (2016). Atypical lateralization of motor circuit functional connectivity in children with autism is associated with motor deficits. *Mol. Autism* 7:35. doi: 10.1186/s13229-016-0096-6

Gaigg, S. B., Cornell, A. S., and Bird, G. (2018). The psychophysiological mechanisms of alexithymia in autism spectrum disorder. *Autism* 22, 227–231. doi: 10.1177/1362361316667062

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/s0896-6273(03)00838-9

Hill, E. L., and Bird, C. M. (2006). Executive processes in Asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia* 44, 2822–2835. doi: 10.1016/j.neuropsychologia.2006.06.007

Hoekstra, R. A., Bartels, M., Cath, D. C., and Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): a study in dutch population and patient groups. *J. Autism Dev. Disord.* 38, 1555–1566. doi: 10.1007/s10803-008-0538-x

Honaga, E., Ishii, R., Kurimoto, R., Canuet, L., Ikezawa, K., Takahashi, H., et al. (2010). Post-movement beta rebound abnormality as indicator of mirror neuron system dysfunction in autistic spectrum disorder: an MEG study. *Neurosci. Lett.* 478, 141–145. doi: 10.1016/j.neulet.2010.05.004

Horn, W. (1983). *LPS Leistungsprüfsystem.* Göttingen: Hogrefe.

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., and Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the autism spectrum quotient (AQ) in a non-clinical sample. *Pers. Individ. Dif.* 43, 1938–1949. doi: 10.1016/j.paid.2007.06.012

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670. doi: 10.1146/annurev.psych.60.110707.163604

Iacoboni, M., and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951. doi: 10.1038/nrn2024

Jansiewicz, E. M., Goldberg, M. C., Newschaffer, C. J., Denckla, M. B., Landa, R., and Mostofsky, S. H. (2006). Motor signs distinguish children with high functioning autism and asperger's syndrome from controls. *J. Autism Dev. Disord.* 36, 613–621. doi: 10.1007/s10803-006-0109-y

Keefer, K. V., Taylor, G. J., Parker, J. D. A., and Bagby, R. M. (2019). Taxometric analysis of the toronto structured interview for alexithymia: further evidence that alexithymia is a dimensional construct. *Assessment* 26, 364–374. doi: 10.1177/1073191117698220

Kemmerer, D. (2014). *Cognitive Neuroscience of Language.* New York, NY: Psychology Press.

Kemmerer, D., Rudrauf, D., Manzel, K., and Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* 48, 826–848. doi: 10.1016/j.cortex.2010.11.001

Kenworthy, L., Yerys, B. E., Anthony, L. G., and Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychol. Rev.* 18, 320–338. doi: 10.1007/s11065-008-9077-7

Kinnaird, E., Stewart, C., and Tchanturia, K. (2019). Investigating alexithymia in autism: a systematic review and meta-analysis. *Eur. Psychiatry* 55, 80–89. doi: 10.1016/j.eurpsy.2018.09.004

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N. V., Chakrabarti, B., Wheelwright, S. J., Auyeung, B., et al. (2012). Cognition in males and females with autism: similarities and differences. *PLoS One* 7:e47198. doi: 10.1371/journal.pone.0047198

Leary, M. R., and Hill, D. A. (1996). Moving on: autism and movement disturbance. *Ment. Retard.* 34, 39–53.

Lehnhardt, F.-G., Falter, C. M., Gawronski, A., Pfeiffer, K., Tepest, R., Franklin, J., et al. (2016). Sex-related cognitive profile in autism spectrum disorders diagnosed late in life: implications for the female autistic phenotype. *J. Autism Dev. Disord.* 46, 139–154. doi: 10.1007/s10803-015-2558-7

Lobben, M., and D'Ascenzo, S. (2015). Grounding grammatical categories: attention bias in hand space influences grammatical congruency judgment of Chinese nominal classifiers. *Front. Psychol.* 6:1299. doi: 10.3389/fpsyg.2015.01299

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., and Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS One* 2:e883. doi: 10.1371/journal.pone.0000883

Lord, C., Rutter, M., DiLavore, P. C., and Risi, S. (2002). *Autism Diagnostic Observation Schedule (ADOS).* Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/bf02172145

MacDonald, M., Lord, C., and Ulrich, D. A. (2013). The relationship of motor skills and social communicative skills in school-aged children with autism spectrum disorder. *Adapt. Phys. Activ. Q.* 30, 271–282. doi: 10.1123/apaq.30.3.271

MacDonald, M., Lord, C., and Ulrich, D. A. (2014). Motor skills and calibrated autism severity in young children with autism spectrum disorder. *Adapt. Phys. Activ. Q.* 31, 95–105. doi: 10.1123/apaq.2013-0068

Mahajan, R., Dirlikov, B., Crocetti, D., and Mostofsky, S. H. (2016). Motor circuit anatomy in children with autism spectrum disorder with or without attention deficit hyperactivity disorder. *Autism Res.* 9, 67–81. doi: 10.1002/aur.1497

McCleery, J. P., Elliott, N. A., Sampanis, D. S., and Stefanidou, C. A. (2013). Motor development and motor resonance difficulties in autism: relevance to early intervention for language and communication skills. *Front. Integr. Neurosci.* 7:30. doi: 10.3389/fnint.2013.00030

Milosavljevic, B., Carter Leno, V., Simonoff, E., Baird, G., Pickles, A., Jones, C. R. G., et al. (2016). Alexithymia in adolescents with autism spectrum disorder: its relationship to internalising difficulties, sensory modulation and social cognition. *J. Autism Dev. Disord.* 46, 1354–1367. doi: 10.1007/s10803-015-2670-8

Ming, X., Brimacombe, M., and Wagner, G. C. (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain Dev.* 29, 565–570. doi: 10.1016/j.braindev.2007.03.002

Mody, M., Shui, A. M., Nowinski, L. A., Golas, S. B., Ferrone, C., O'Rourke, J. A., et al. (2017). Communication deficits and the motor system: exploring patterns of associations in autism spectrum disorder (ASD). *J. Autism Dev. Disord.* 47, 155–162. doi: 10.1007/s10803-016-2934-y

Moseley, R., Carota, F., Hauk, O., Mohr, B., and Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cereb. Cortex* 22, 1634–1647. doi: 10.1093/cercor/bhr238

Moseley, R. L., Mohr, B., Lombardo, M. V., Baron-Cohen, S., Hauk, O., and Pulvermüller, F. (2013). Brain and behavioral correlates of action semantic deficits in autism. *Front. Hum. Neurosci.* 7:725. doi: 10.3389/fnhum.2013.00725

Moseley, R. L., and Pulvermüller, F. (2014). Nouns, verbs, objects, actions, and abstractions: local fMRI activity indexes semantics, not lexical categories. *Brain Lang.* 132, 28–42. doi: 10.1016/j.bandl.2014.03.001

Moseley, R. L., and Pulvermüller, F. (2018). What can autism teach us about the role of sensorimotor systems in higher cognition? New clues from studies on language, action semantics, and abstract emotional concept processing. *Cortex* 100, 149–190. doi: 10.1016/j.cortex.2017.11.019

Moseley, R. L., Pulvermüller, F., Mohr, B., Lombardo, M. V., Baron-Cohen, S., and Shtyrov, Y. (2014). Brain routes for reading in adults with and without autism: EMEG evidence. *J. Autism Dev. Disord.* 44, 137–153. doi: 10.1007/s10803-013-1858-z

Moseley, R. L., Shtyrov, Y., Mohr, B., Lombardo, M. V., Baron-Cohen, S., and Pulvermüller, F. (2015). Lost for emotion words: what motor and limbic brain activity reveals about autism and semantic theory. *Neuroimage* 104, 413–422. doi: 10.1016/j.neuroimage.2014.09.046

Mostofsky, S. H., Burgess, M. P., and Gidley Larson, J. C. (2007). Increased motor cortex white matter volume predicts motor impairment in autism. *Brain* 130, 2117–2122. doi: 10.1093/brain/awm129

Mostofsky, S. H., Powell, S. K., Simmonds, D. J., Goldberg, M. C., Caffo, B., and Pekar, J. J. (2009). Decreased connectivity and cerebellar activity in autism during motor task performance. *Brain* 132, 2413–2425. doi: 10.1093/brain/awp088

Nishitani, N., Avikainen, S., and Hari, R. (2004). Abnormal imitation-related cortical activation sequences in Asperger's syndrome. *Ann. Neurol.* 55, 558–562. doi: 10.1002/ana.20031

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cogn. Brain Res.* 24, 190–198. doi: 10.1016/j.cogbrainres.2005.01.014

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Papadopoulos, N., McGinley, J., Tonge, B., Bradshaw, J., Saunders, K., Murphy, A., et al. (2012). Motor proficiency and emotional/behavioural disturbance in autism and Asperger's disorder: another piece of the neurological puzzle? *Autism* 16, 627–640. doi: 10.1177/1362361311418692

Popp, M., Trumpp, N. M., and Kiefer, M. (2016). Feature-specific event-related potential effects to action- and sound-related verbs during visual word recognition. *Front. Hum. Neurosci.* 10:637. doi: 10.3389/fnhum.2016.00637

Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci.* 22, 253–279. doi: 10.1017/S0140525X9900182X

Pulvermüller, F. (2012). Meaning and the brain: the neurosemantics of referential, interactive, and combinatorial knowledge. *J. Neurolinguistics* 25, 423–459. doi: 10.1016/j.jneuroling.2011.03.004

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797. doi: 10.1111/j.1460-9568.2005.03900.x

Pulvermüller, F., Moseley, R. L., Egorova, N., Shebani, Z., and Boulenger, V. (2014). Motor cognition-motor semantics: action perception theory of cognition and communication. *Neuropsychologia* 55, 71–84. doi: 10.1016/j.neuropsychologia.2013.12.002

Rizzolatti, G., and Fabbri-Destro, M. (2010). Mirror neurons: from discovery to autism. *Exp. Brain Res.* 200, 223–237. doi: 10.1007/s00221-009-2002-3

Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274. doi: 10.1038/nrn2805

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., et al. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol. Autism* 6:2. doi: 10.1186/2040-2392-6-2

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., et al. (2016). Subgrouping siblings of people with autism: identifying the broader autism phenotype. *Autism Res.* 9, 658–665. doi: 10.1002/aur.1544

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., and Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. *Soc. Neurosci.* 3, 97–112. doi: 10.1080/17470910701577020

Stevenson, J. L., and Hart, K. R. (2017). Psychometric properties of the autism-spectrum quotient for assessing low and high levels of autistic traits in college students. *J. Autism Dev. Disord.* 47, 1838–1853. doi: 10.1007/s10803-017-3109-1

Stevenson, J. L., Lindley, C. E., and Murlo, N. (2017). Retrospectively assessed early motor and current pragmatic language skills in autistic and neurotypical children. *Percept. Mot. Skills* 124, 777–794. doi: 10.1177/0031512517710379

Taylor, G. J., Ryan, D., and Bagby, M. (1985). Toward the development of a new self-report alexithymia scale. *Psychother. Psychosom.* 44, 191–199. doi: 10.1159/000287912

Thompson, A., Murphy, D., Dell'Acqua, F., Ecker, C., McAlonan, G., Howells, H., et al. (2017). Impaired communication between the motor and somatosensory homunculus is associated with poor manual dexterity in autism spectrum disorder. *Biol. Psychiatry* 81, 211–219. doi: 10.1016/j.biopsych.2016.06.020

Tiffin, J., and Asher, E. J. (1948). The Purdue Pegboard: norms and studies of reliability and validity. *J. Appl. Psychol.* 32, 234–247. doi: 10.1037/h0061266

Tombaugh, T. (2004). Trail making test A and B: normative data stratified by age and education. *Arch. Clin. Neuropsychol.* 19, 203–214. doi: 10.1016/S0887-6177(03)00039-8

Travers, B. G., Bigler, E. D., Tromp do, P. M., Adluru, N., Destiche, D., Samsin, D., et al. (2015). Brainstem white matter predicts individual differences in manual motor difficulties and symptom severity in autism. *J. Autism Dev. Disord.* 45, 3030–3040. doi: 10.1007/s10803-015-2467-9

Travers, B. G., Powell, P. S., Klinger, L. G., and Klinger, M. R. (2013). Motor difficulties in autism spectrum disorder: linking symptom severity and postural stability. *J. Autism Dev. Disord.* 43, 1568–1583. doi: 10.1007/s10803-012-1702-x

Uljarević, M., Hedley, D., Alvares, G. A., Varcin, K. J., and Whitehouse, A. J. O. (2017). Relationship between early motor milestones and severity of restricted and repetitive behaviors in children and adolescents with autism spectrum disorder. *Autism Res.* 10, 1163–1168. doi: 10.1002/aur.1763

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., and Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci. Biobehav. Rev.* 35, 407–426. doi: 10.1016/j.neubiorev.2010.04.007

Vonk, J. M. J., Obler, L. K., and Jonkers, R. (2019). Levels of abstractness in semantic noun and verb processing: the role of sensory-perceptual and sensory-motor information. *J. Psycholinguist. Res.* 48, 601–615. doi: 10.1007/s10936-018-9621-4

Wadsworth, H. M., Maximo, J. O., Lemelman, A. R., Clayton, K., Sivaraman, S., Deshpande, H. D., et al. (2017). The Action Imitation network and motor imitation in children and adolescents with autism. *Neuroscience* 343, 147–156. doi: 10.1016/j.neuroscience.2016.12.001

Wallace, G. L., Kenworthy, L., Pugliese, C. E., Popal, H. S., White, E. I., Brodsky, E., et al. (2016). Real-world executive functions in adults with autism spectrum disorder: profiles of impairment and associations with adaptive functioning and co-morbid anxiety and depression. *J. Autism Dev. Disord.* 46, 1071–1083. doi: 10.1007/s10803-015-2655-7

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., et al. (2006). Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ). *Brain Res.* 1079, 47–56. doi: 10.1016/j.brainres.2006.01.012

Wing, L., and Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *J. Autism Dev. Disord.* 9, 11–29. doi: 10.1007/BF01531288

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., and Baron-Cohen, S. (2005). Screening adults for asperger syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. *J. Autism Dev. Disord.* 35, 331–335. doi: 10.1007/s10803-005-3300-7

Yang, H., Lin, Q., Han, Z., Li, H., Song, L., Chen, L., et al. (2017). Dissociable intrinsic functional networks support noun-object and verb-action processing. *Brain Lang.* 175, 29–41. doi: 10.1016/j.bandl.2017.08.009

Yudes, C., Dominguez, A., Cuetos, F., and De Vega, M. (2016). The time-course of processing of grammatical class and semantic attributes of words: dissociation by means of ERP. *Psiocologica* 37, 105–126.

Zhao, B., Dang, J., and Zhang, G. (2017). EEG source reconstruction evidence for the noun-verb neural dissociation along semantic dimensions. *Neuroscience* 359, 183–195. doi: 10.1016/j.neuroscience.2017.07.019

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 04 October 2019
doi: 10.3389/fnhum.2019.00344

Auditory and Somatosensory Interaction in Speech Perception in Children and Adults

Paméla Trudeau-Fisette [1,2]****, Takayuki Ito*** [3,4] ***and Lucie Ménard*** [1,2]

[1]Laboratoire de Phonétique, Université du Québec à Montréal, Montreal, QC, Canada, [2]Centre for Research on Brain, Language and Music, Montreal, QC, Canada, [3]GIPSA-Lab, CNRS, Grenoble INP, Université Grenoble Alpes, Grenoble, France, [4]Haskins Laboratories, Yale University, New Haven, CT, United States

Multisensory integration (MSI) allows us to link sensory cues from multiple sources and plays a crucial role in speech development. However, it is not clear whether humans have an innate ability or whether repeated sensory input while the brain is maturing leads to efficient integration of sensory information in speech. We investigated the integration of auditory and somatosensory information in speech processing in a bimodal perceptual task in 15 young adults (age 19–30) and 14 children (age 5–6). The participants were asked to identify if the perceived target was the sound /e/ or /ø/. Half of the stimuli were presented under a unimodal condition with only auditory input. The other stimuli were presented under a bimodal condition with both auditory input and somatosensory input consisting of facial skin stretches provided by a robotic device, which mimics the articulation of the vowel /e/. The results indicate that the effect of somatosensory information on sound categorization was larger in adults than in children. This suggests that integration of auditory and somatosensory information evolves throughout the course of development.

Keywords: multisensory integration, speech perception, auditory and somatosensory feedback, adults, children, categorization, maturation

Edited by:
Carmen Moret-Tatay,
Catholic University of Valencia San Vicente Mártir, Spain

Reviewed by:
Camila Rosa De Oliveira,
Faculdade Meridional (IMED), Brazil
Nuria Senent-Capuz,
Catholic University of Valencia San Vicente Mártir, Spain

****Correspondence:***
Paméla Trudeau-Fisette
ptrudeaufisette@gmail.com

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *21 March 2019*
Accepted: *18 September 2019*
Published: *04 October 2019*

Citation:
Trudeau-Fisette P, Ito T and Ménard L (2019) Auditory and Somatosensory Interaction in Speech Perception in Children and Adults. Front. Hum. Neurosci. 13:344. doi: 10.3389/fnhum.2019.00344

INTRODUCTION

From our first day of life, we are confronted with multiple sensory inputs such as tastes, smells, and touches. Unconsciously, related inputs are combined into a single input with rich information. Multisensory integration (MSI), also called multimodal integration, is the ability of the brain to assimilate cues from multiple sensory modalities that allows us to benefit from the information from each sense to reduce perceptual ambiguity and ultimately reinforce our perception of the world (Stein and Meredith, 1993; Stein et al., 1996; Robert-Ribes et al., 1998; Molholm et al., 2002). MSI holds a prominent place in the way that information is processed, by shaping how inputs are perceived. This merging of various sensory inputs into common neurons was typically assumed to occur late in the perceptual process stream (Massaro, 1999), but recent studies in neurophysiology have even demonstrated that MSI can occur in the early stages of cortical processing, even in brain regions typically associated with lower-level processing of uni-sensory inputs (Macaluso et al., 2000; Foxe et al., 2002; Molholm et al., 2002; Mishra et al., 2007; Raij et al., 2010; Mercier et al., 2013).

While some researchers have suggested that an infant's brain is likely equipped with multisensorial functionality at birth (Bower et al., 1970; Streri and Gentaz, 2004), others have suggested that MSI likely develops over time as a result of experiences (Birch and Lefford, 1963;

Yu et al., 2010; Burr and Gori, 2011). Several studies support the latter hypothesis. For example, studies have demonstrated that distinct sensory systems develop at different rates and in different ways, which suggests that several mechanisms are implicated in MSI depending on the type of interactions (Walker-Andrews, 1994; Gori et al., 2008; Burr and Gori, 2011; Dionne-Dostie et al., 2015). For example, researchers have reported that eye-hand coordination, a form of somatovisual interaction, can be observed in infants as young as a week old (Bower et al., 1970), and audiovisual association of phonetic information emerges around 2 months of age (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003), but audiovisual integration in spatial localization behavior does not appear before 8 months of age (Neil et al., 2006).

Ultimately, although it is still unclear whether an innate system enables MSI in humans, data from infants, children, and adults suggest that unimodal and multimodal sensory experiences and brain maturation enables the establishment of efficient integration processing (Rentschler et al., 2004; Krakauer et al., 2006; Neil et al., 2006; Gori et al., 2008; Nardini et al., 2008; Hillock et al., 2011; Stein et al., 2014) and that multisensory tasks in school-aged and younger children are executed through unimodal dominance rather than integration abilities (McGurk and Power, 1980; Hatwell, 1987; Misceo et al., 1999; Burr and Gori, 2011). Moreover, according to the intersensory redundancy hypothesis, perception of multimodal information is only facilitated when information from various sources is redundant, and not when the information is conflicting (Bahrick and Lickliter, 2000, 2012).

Multimodal integration is crucial for speech development. According to the associative view, during infancy, the acoustic features of produced and perceived speech are associated with felt and seen articulatory movements required for their production (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003; Pons et al., 2009; Yeung and Werker, 2013). Once acoustic information and proprioceptive feedback information are strongly linked together, this becomes part of an internal multimodal speech model (Guenther and Perkell, 2004; Tourville and Guenther, 2011; Guenther and Vladusich, 2012).

MSI can sometimes be overlooked in speech perception since speakers frequently have one dominant sensory modality (Hecht and Reiner, 2009; Lametti et al., 2012). However, even though audition is the dominant type of sensory information in speech perception, many researchers have suggested that other sensory modalities also play a role in speech processing (Perrier, 1995; Tremblay et al., 2003; Skipper et al., 2007; Ito et al., 2009; Lametti et al., 2012). The McGurk effect, a classic perceptual illusion resulting from incongruent simultaneous auditory and visual cues about consonants clearly demonstrates that information from multiple sensory channels is unconsciously integrated during speech processing (McGurk and MacDonald, 1976).

In the current study, we examined the integration of auditory and somatosensory interaction in speech perception. Previous research has suggested that to better understand how different types of sensory feedback interact in speech perception, we need to better understand how and when this becomes mature.

Hearing is one of the first sensory modalities to emerge in humans. While still *in utero*, babies can differentiate speech from non-speech and distinguish variability in speech length and intensity (for a review on auditory perception in the fetus, see Lecanuet et al., 1995). After birth, babies are very soon responsive to various rhythmic and intonation sounds (Demany et al., 1977) and can distinguish phonemic features such as voicing, manner, and place of articulation (Eimas et al., 1971). Specific perceptual aspects of one's first language, such as sensitivity to phonemes and phonotactic properties, are refined by the first year of life (Kuhl, 1991). Although auditory abilities become well established in the early years of life, anatomical changes and experiences will guide the development of auditory skills throughout childhood (Arabin, 2002; Turgeon, 2011).

Little is known about the development of oral somatosensory abilities in typically developing children. Yet, some authors have worked on the development of oral stereognosis in children and adults, where stereognosis is the ability to perceive and recognize the form of an object in the absence of visual and auditory information, by using tactile information. In oral stereognosis, the form of an object is recognized by exploring tactile information such as texture, size or spatial properties, in the oral cavity. This is usually evaluated by comparing the ability of children and adults to differentiate or identify small plastic objects in their mouths. Researchers have reported that oral sensory discrimination skills depend on age (McDonald and Aungst, 1967; Dette and Linke, 1982; Gisel and Schwob, 1988). McDonald and Aungst (1967) showed that 6- to 8-year-old children correctly matched half of the presented forms; 17- to 31-year-old adolescents and adults had perfect scores; and scores declined significantly with age among the 52- to 89-year-olds. Dette and Linke (1982) found similar results in 3- to 17-year-olds. The effect of age was also found in younger vs. older children. Kumin et al. (1984) showed that among 4- to 11-year-olds, the older children had significantly better oral stereognosis scores than younger children. Gisel and Schwob (1988) reported that 7- and 8-year-old children had better identification skills in an oral stereognosis experiment than 5- and 6-year-old children. Interestingly, only the 8-year-old children showed a learning effect, in that they got better scores as the experiment progressed.

To explain this age-related improvement in oral stereognosis, it was suggested that oral stereognosis maturity is achieved when the growth of the oral and facial structures is complete (McDonald and Aungst, 1967; Gisel and Schwob, 1988). This explanation is consistent with vocal tract growth data that shows that while major changes occur in the first 3 years of life (Vorperian et al., 1999), important growth of the pharyngeal region is observed between puberty and adulthood (Fitch and Giedd, 1999) and multidimensional maturity of the vocal tract is not reached until adulthood (Boë et al., 2007, 2008).

A few recent studies have suggested that there is a link between auditory and somatosensory information in multimodal integration.

Lametti et al. (2012) proposed that sensory preferences in the specification of speech motor goals could mediate responses to real-time manipulations, which would explain the important variability in compensatory behavior to an auditory manipulation (Purcell and Munhall, 2006; Villacorta et al., 2007; MacDonald et al., 2010). They point out that one's own auditory feedback is not the only reliable source of speech monitoring and, in line with the internal speech model theory, that somatosensory feedback would also be considered in speech motor control. In agreement with this concept, Katseff et al. (2012) suggested that partial compensation in auditory manipulation of real-time speech could be because both auditory and somatosensory feedback system monitor speech motor control and therefore, the two systems are competing when large sensory manipulation affects only one of the sensory channels.

A recent study of speech auditory feedback perturbations in blind and sighted speakers supports the latter explanation. It showed that typically developing adults, whose somatosensory goals are narrowed by vision were more likely to tolerate large discrepancies between the expected and produced auditory outcome, whereas blind speakers, whose auditory goals had primacy over somatosensory ones, tolerated larger discrepancies between their expected and produced somatosensory feedback. In this sense, blind speakers were more inclined to adopt unusual articulatory positions to minimize divergences of their auditory goals (Trudeau-Fisette et al., 2017).

Researchers have also suggested that acoustic and somatosensory cues are integrated. As far as we know, Von Schiller (cited in Krueger, 1970; Jousmäki and Hari, 1998) was the first one to report that sound could modulate touch. Indeed, although he was mainly focused on the interaction between auditory and visual cues, he showed in his 1932s article that auditory stimuli, such as tones and noise bursts, could influence an object's physical perception. Since then, studies have shown how manipulations of acoustic frequencies or even changes in their prevalence can influence the tactile perception of objects, events, and skin deformation such as their perceived smoothness, occurrence, or magnitude (Krueger, 1970; Jousmäki and Hari, 1998; Guest et al., 2002; Hötting and Röder, 2004; Ito and Ostry, 2010). Multimodal integration was stronger when both perceptual sources were presented simultaneously (Jousmäki and Hari, 1998; Guest et al., 2002).

This interaction between auditory and tactile channels is also found in the opposite direction, in that somatosensory inputs can influence the perception of sounds. For example, Schürmann et al. (2004) showed that vibrotactile cues can influence the perception of sound loudness. Later, Gick and Derrick (2009) demonstrated that aerotactile inputs could modulate the perception of a consonant's oral property.

Somatosensory information coming from orofacial areas is somewhat different from those typically intended. Kinesthetic feedback usually refers to information retrieved from position, movement, and receptors in muscles and articulators (Proske and Gandevia, 2009). However, some of the orofacial regions involved in speech production movement are devoid of muscle proprioceptors. Therefore, the somatosensory information guiding our perception and production abilities likely also come from cutaneous mechanoreceptors (Johansson et al., 1988; Ito and Gomi, 2007; Ito and Ostry, 2010).

Although many studies have reported on the role of somatosensory information derived from orofacial movement in speech production (Tremblay et al., 2003; Nasir and Ostry, 2006; Ito and Ostry, 2010; Feng et al., 2011; Lametti et al., 2012), few studies have reported its role in speech perception.

Researchers recently investigated the contribution of somatosensory information on speech perception mechanisms. Ito et al. (2009) designed a bimodal perceptual task experiment where they asked participants to identify if the perceived target was the word "head" or "had." When the acoustic targets (all members of the "head/had" continuum) were perceived simultaneously to a skin manipulation recalling the oral articulatory gestures implicated in the production of the vowel /ϵ/, the identification rate of the target "head" was significantly improved. The researchers also tested different directions of the orofacial muscle manipulation and established that the observed effect was only found if the physical manipulation reflected a movement required in speech production (Ito et al., 2009).

Somatosensory information appears to even be involved in the processing of higher-level perceptual concepts (Ogane et al., 2017). In a similar perceptual task, participants were asked to identify if the perceived acoustic target was "l'affiche" (the poster) or "la fiche" (the form). The authors showed that the appropriate temporal positions of somatosensory skin manipulation in the stimulus word, simulating somatosensory inputs concerning the hyperarticulation of either the vowel /a/ or the vowel /i/, could affect the categorization of the lexical target.

Although further study would reinforce these findings, these experiments highlight the fact that the perception of linguistic inputs can be influenced by the manipulation of cutaneous receptors involved in speech motion (Ito et al., 2009, 2014; Ito and Ostry, 2010), and furthermore, attest of a strong link between auditory and somatosensory channels within the multimodal aspect of speech perception in adults.

The fact that sounds discrimination if facilitated when included in the infants' babbling register (Vihman, 1996) is surely part of the growing body of evidence that demonstrates how somatosensory information that is derived from speech movement also influences speech perception in young speakers (DePaolis et al., 2011; Bruderer et al., 2015; Werker, 2018). However, to our knowledge, only two studies have investigated how somatosensory feedback is involved in speech perception abilities in children (Yeung and Werker, 2013; Bruderer et al., 2015). In both studies, the researchers manipulated oral somatosensory feedback by constraining tongue or lip movement, thus forcing the adoption of a precise articulatory position. Although MSI continues to evolve until late childhood (Ross et al., 2011), these two experiments in toddlers shed light on how this phenomenon emerges.

In their 2013 article, Yeung and Werker (2013) reported that when 4- and 5-month-old infants were confronted with incongruent auditory and labial somatosensory cues, they were more likely to fix the visual demonstration corresponding to the vowel perceived through the auditory channel. In contrast, congruent auditory and somatosensory cues did not call for the need to add a corresponding visual representation of the perceived vowel.

Also using a looking-time procedure, Bruderer et al. (2015) focused on the role of language experience on the integration of somatosensory information. They found that the ability of 6-month-old infants to discriminate between the non-native dental /d̪/ and the retroflex /ɖ/ Hindi consonant was influenced by the insertion of a teething toy. When the toddlers' tongue movements were restrained, they showed no evidence of phonetic contrast discrimination of tongue tip position. As shown by Ito et al. (2009), the effect of somatosensory cues was only observed if the perturbed articulator would have been involved in the production of the sound that was heard.

While these two studies mainly focused on perceptual discrimination rather than categorical representation of speech, they suggest that proprioceptive information resulting from static articulatory perturbation plays an important role in speech perception mechanisms in toddlers and that the phenomenon of multimodal integration in the perception-production speech model starts early in life. The authors suggested that, even at a very young age, babies can recognize that information can come from multiple sources and they react differently when the sensory sources are compatible. However, it is still unknown when children begin to integrate various sensory sources to treat them as a single sensory source.

In the current study, we aimed to investigate how dynamic somatosensory information from orofacial cutaneous receptors is integrated in speech processing in children compared to adults. Based on previous research, we hypothesized that: (1) when somatosensory inputs are presented simultaneously with auditory inputs, this affects their phonemic categorization; (2) auditory and somatosensory integration is stronger in adults than in children; and (3) MSI is facilitated when both types of sensory feedback are consistent.

MATERIALS AND METHODS

Participants

We recruited 15 young adults (aged 19–30), including eight females. We also recruited 21 children (aged 4–6) and after excluding seven children due to equipment malfunction (1), non-completion (2), or inability to understand the task (4), this left 14 children (aged 5–6) including 10 females, for the data analysis. Five- to six-year-old is a particularly interesting age window since children master all phonemes of their native language. However, they have not yet entered the fluent reading stage, during which explicit teaching of reading has been shown to alter multimodal perceptual (Horlyck et al., 2012).

All participants were native speakers of Canadian French and were tested for pure-tone detection threshold using an adaptive method (DT < 25 dB HL at 250, 500, 1,000, 2,000, 4,000 and 8,000 Hz). None of the participants reported having speech or language impairments. The research protocol was approved by the Université du Québec à Montréal's Institutional Review Board (no 2015-05-4.2) and all participants (or the children's parents) gave written informed consent. The number of participants was limited due to the age of the children and the length of the task (3 different tasks were executed on the same day).

Experimental Procedure

As in the task used by Ito et al. (2009), the participants were asked to identify the vowel they perceived and were asked to choose between /e/ and /ø/. Based on Ménard and Boe (2004), the auditory stimulus consisted of 10 members of a synthesized /e–ø/ continuum generated using the Maeda model (see **Table 1**). This continuum was created such that the first four formants were equally distributed from those corresponding to the natural endpoint tokens of /e/ and /ø/. To ensure that the children understood the difference between the two vocalic choices, the vowel /e/ was represented by an image of a fairy (/e/ as in fée) and the vowel /ø/ was represented by an image of a fire (/ø/ as in feu). Since, we wanted to minimize large head movements during the experiment, the children were asked to point out the image corresponding to their answers. Both images were placed in front of them at shoulder level, three feet away from each other on the horizontal plane. The adults were able to use the keyboard without looking at it and they used the right and left arrows to indicate their responses.

Figure 1 shows the experimental set-up for the facial skin stretch perturbations. The participants were seated with their backs to a Phantom 1.0 device (SensAble Technologies) and they wore headphones (Sennheiser HD 380 pro). This small unit, composed of a robotic arm to which a wire is attached, allows for minor lateral skin manipulation at the side of the mouth, where small plastic tabs (2 mm × 3 mm), located on the ends of the wire, were placed with double-sided tape. The robotic arm was programed to ensure that when a four Newton

TABLE 1 | Formant and bandwidth values of the synthesized stimuli used in the perceptual task.

	Formant values					Bandwidths values				
	F1	F2	F3	F4	F5	B1	B2	B3	B4	B5
Auditory stimuli										
1	364	1,922	2,509	3,550	4,000	48	55	60	50	100
2	364	1,892	2,469	3,500	4,000	48	55	60	50	100
3	364	1,862	2,429	3,450	4,000	48	55	60	50	100
4	364	1,832	2,389	3,400	4,000	48	55	60	50	100
5	364	1,802	2,349	3,350	4,000	48	55	60	50	100
6	364	1,772	2,309	3,300	4,000	48	55	60	50	100
7	364	1,742	2,269	3,250	4,000	48	55	60	50	100
8	364	1,712	2,229	3,200	4,000	48	55	60	50	100
9	364	1,682	2,189	3,150	4,000	48	55	60	50	100
10	364	1,652	2,149	3,100	4, 000	48	55	60	50	100

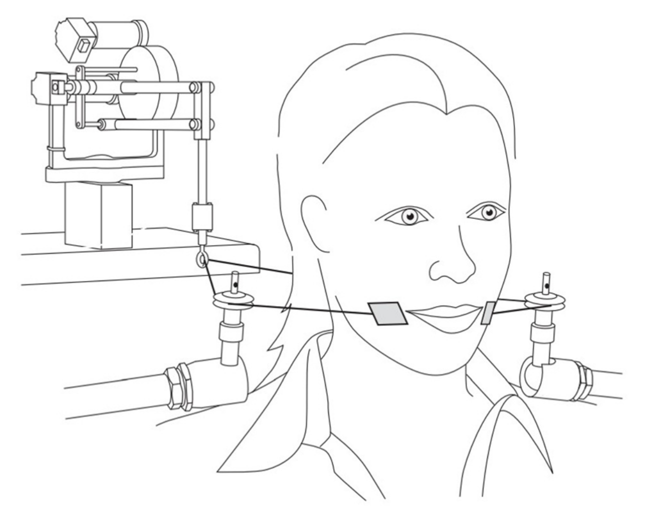


FIGURE 1 | Experimental set up for facial skin stretch perturbations (reproduced with permission from Ito and Ostry, 2010).

flexion force was administered it led to a 10- to 15-mm lateral skin stretch.

When this facial skin stretch is applied at lateral to the oral angle in the backward direction as shown in the figure, it mimics the articulation associated with the production of the unrounded vowel /e/. Therefore, auditory and somatosensory feedback was either congruent (with /e/-like auditory inputs) or incongruent (with /ø/-like auditory inputs). As stated early, cutaneous receptors found in the within the labial area provides speech related kinesthetic information (Ito and Gomi, 2007). Since the skin manipulation was programed to be perceived at the same time as the auditory stimuli, it was possible to investigate the contribution of the somatosensory system to the perceptual processing of the speech targets.

The auditory stimuli were presented in 20 blocks of 10 trials each. Within each block, all members of the 10-step continuum were presented in a random order. For half of the trials, only the auditory stimulus was presented (unimodal condition). For the other half of the trials, a facial skin manipulation was also applied (bimodal condition). Alternate blocks of unimodal and bimodal conditions were presented to the participants. In total, 200 perceptual judgments were collected, 100 in the auditory-only condition and 100 in the combined auditory and skin-stretch condition.

Data Analysis

For each participant, stimulus, and condition, we calculated the percentage of /e/ responses. The experiment was closely monitored, and the responses in trials where a short pause was requested by the participant were excluded from the analysis. In doing so, we sought to eliminate categorical judgments for which the participants were no longer in a position to properly respond to the task (fewer than 1.1% and 0.2% of all responses were excluded for children and adults, respectively). These perceptual scores were then fitted onto a logistic regression model (Probit model) to obtain psychometric functions from which the labeling slopes and 50% crossover boundaries were computed. The value of the slope corresponds to the sharpness of the categorization (the lower the value, the more distinct the categorization), while the boundary value indicates the location of the categorical boundary between the two vowel targets (the higher the value, the more toward /ø/ the frontier). Using the lme4 package in R, we carried out a linear mixed-effects model (Baayen et al., 2008) for both the steepness of the slopes and the category boundaries in which group (adult or children) and condition (unimodal or bimodal) were specified as fixed factors and individual participant was defined as a random factor.

Each given answer (5,800 perceptual judgments collected from 29 participants) was fitted into a linear mixed-effects model where fixed factors included stimuli (the 10-step continuum), group (adult or children), and condition (unimodal or bimodal), and the random factor was the individual participant. The mean categorization of the first and last two stimuli was also compared. Once again, the averages of the given answers (116 mean perceptual judgments collected from 29 participants) were fitted into a linear mixed-effects model where the fixed variables included stimuli (head stimuli or tail stimuli), group (adult or children), and condition (unimodal or bimodal) and where the random variable was the individual participant. Finally, independent *t*-tests were carried out in order to compare variability in responses between both experimental groups and conditions. In both cases, Kolmogorov–Smirnov tests indicated that categorizations followed a normal distribution.

RESULTS

The overall percentage of /e/ responses for each stimulus is shown in **Figure 2**. The data were averaged across speakers, within both groups. **Figure 3** displays the values for the labeling slope (distinctiveness of the vowels' categorization) and 50% crossover boundary (location of the categorical frontier) averaged across experimental conditions and groups. As can be seen in both figures, regardless of the experimental condition, the children had greater variations in overall responses compared to the adults, which was confirmed in an independent *t*-test ($t_{(38)} = 2.792, p < 0.01$).

Psychometric Functions

Labeling Slope Results

The linear mixed-effects model revealed a significant main effect of group on the steepness of the slope ($\chi^2_{(1)} = 23.549, p < 0.001$), indicating that there was more categorical perception in adults than in children (see **Figure 2**, black lines and **Figure 3**, left-hand part of the graph).

Although no effect of condition as a main effect was observed ($\chi^2_{(1)} = 3.618, p > 0.05$), a significant interaction between group and condition was found ($\chi^2_{(1)} = 4.956, p < 0.05$). *Post hoc* analysis revealed that in the bimodal condition the slope of the labeling function was more abrupt for the adults ($z = -3.153, p < 0.01$) but not for the children, suggesting that the skin stretch condition led to a more categorical identification of the stimuli in adults only.

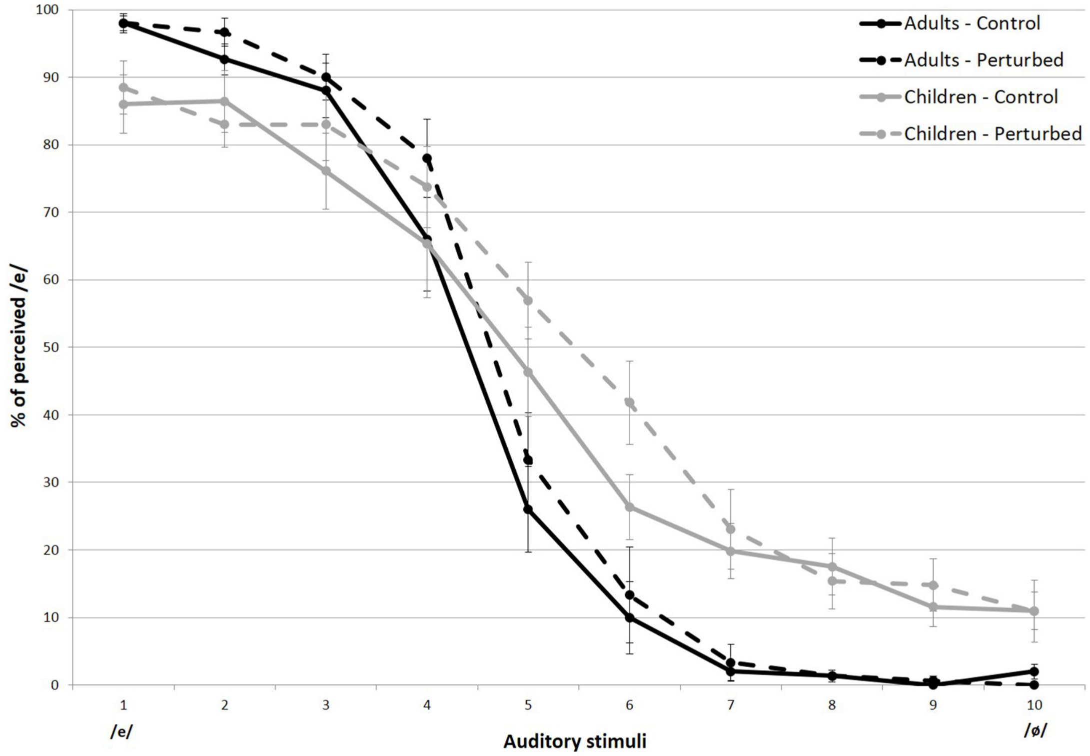


FIGURE 2 | Percent identification of the vowel [e] for stimuli on the [e–ø] continuum, in both experimental conditions, for both groups. Error bars indicate standard errors.

The 50% Crossover Boundary Results

A linear mixed-effects model analysis carried out on the 50% crossover boundaries revealed a single main effect of condition ($\chi^2_{(1)} = 9.245$, $p < 0.01$). For both groups, the skin stretch perturbation led to a displacement of the 50% crossover boundary. In the bimodal condition (A+SS), the boundary was located closer to /ø/ than in the unimodal condition (A). This result is consistent with the expected effect of the skin stretch perturbation; more stimuli were perceived as /e/ than /ø/. No effect of group, as a main effect or with condition was found. The results are presented in **Figure 2** and in **Figure 3**, in the right-hand part of the graphs.

Categorical Judgments

A linear mixed-effects model analysis performed on the categorical judgments revealed that in addition to the expected main effect of stimuli ($\chi^2_{(1)} = 3652.4$, $p < 0.001$), there were significant effects of group ($\chi^2_{(1)} = 4.586$, $p < 0.05$) and condition ($\chi^2_{(1)} = 15.736$, $p < 0.001$), suggesting that children and adults did not categorize the stimuli in a similar manner and that both experimental conditions prompted different categorization. Moreover, a significant interaction of group and stimuli ($\chi^2_{(1)} = 144.52$, $p < 0.001$) revealed that irrespectively of the experimental condition, some auditory stimuli were categorized differently by the two groups.

Post hoc tests revealed that whether a skin stretch manipulation was applied or not, stimulus 7 (A $z = -3.795$, $p < 0.1$ A+SS $z = -4.648$, $p < 0.01$), 8 (A $z = -3.445$, $p < 0.5$ A+SS $z = -3.544$, $p < 0.1$) and 9 (A $z = -3.179$, $p < 0.5$ A+SS $z = -4.347$, $p < 0.01$) were more systematically identified as /ø/ by the adults than by the children. While no other two-way interactions were found, a significant three-way interaction of group, condition, and stimuli was observed ($\chi^2_{(4)} = 117.26$, $p < 0.001$) suggesting that, for some specific stimuli, the skin stretch condition affected the perceptual judgment of both groups in a different manner.

First, it was found that the skin stretch manipulation had a greater effect on stimulus 6, in children only ($z = -3.251$, $p < 0.5$). For this group, the skin stretch condition caused a 15.8% increase of /e/ labeling on stimulus 6. For the adults, the addition of somatosensory cues only led to a 3.3% increase in /e/ categorization.

Although less expected, the skin stretch manipulation also led to some perceptual changes at the endpoint of the auditory continuum. As shown in **Figure 2**, stimulus 2 ($z = 3.053$, $p < 0.5$) and stimulus10 ($z = -3.734$, $p < 0.1$) were labeled differently by the two groups, but only in the bimodal condition. In fact, stimulus 2 (an /e/-like stimulus) was more likely to be identified as an /e/ by the adults in the experimental condition. In contrast, children were less inclined to label it so. As for stimulus 10 (an /ø/-like stimulus), the addition of somatosensory inputs

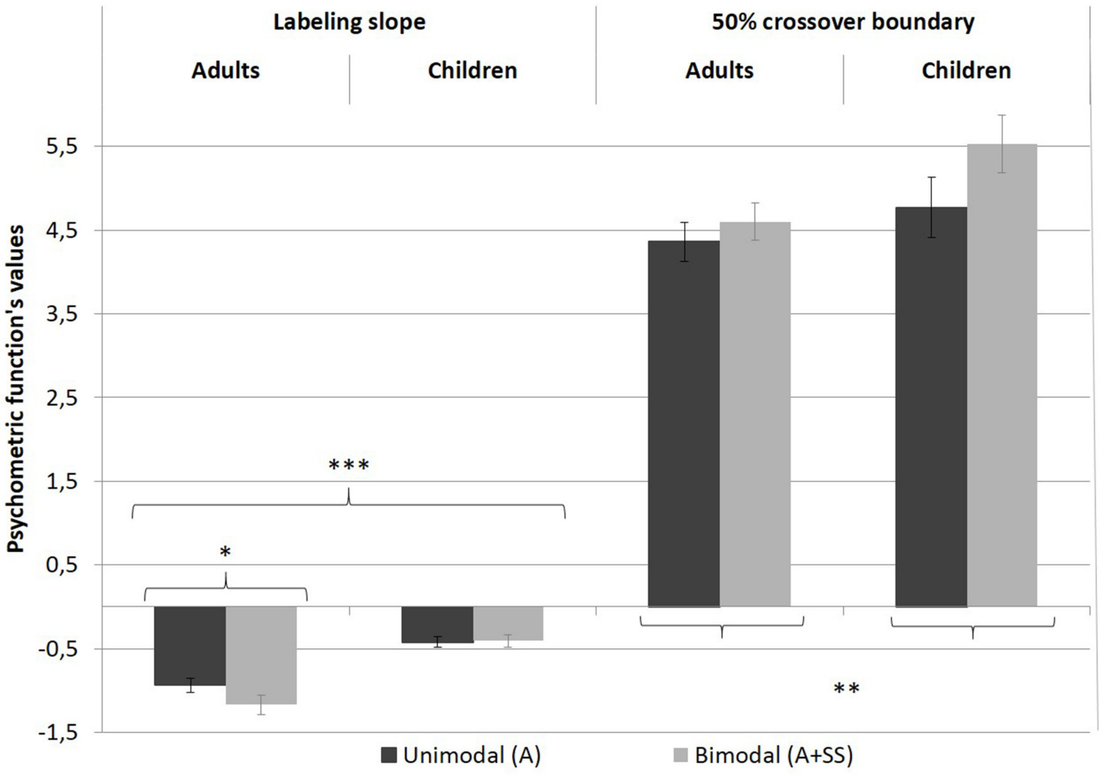


FIGURE 3 | Psychometric functions of labeling slope and 50% crossover boundary, in both experimental conditions, for both groups. Error bars indicate standard errors. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

decreased the correct identification rate in children only. In adults, although it barely affected their categorical judgments, the skin stretch manipulation mimicking the articulatory gestures of the vowel /e/ resulted in an increase of /ø/ labeling, as if it had a reverse effect.

Last, a comparison of mean categorizations of the first and last two stimuli revealed a main effect of stimuli ($\chi^2_{(1)} = 313.52$, $p < 0.001$) and a significant interaction of group and stimuli ($\chi^2_{(1)} = 36.260$, $p < 0.001$). More importantly, it also revealed a 3-way interaction of group, condition, and stimuli ($\chi^2_{(4)} = 37.474$, $p < 0.001$). *Post hoc* tests indicated that those endpoint stimuli of the continuum were categorized differently by the two groups, but only when a skin stretch manipulation was applied. In agreement with previous results, in the skin stretch condition, children labeled more /e/-like stimuli as /ø/ ($z = 3.434$, $p < 0.5$), and more /ø/-like stimuli as /e/ ($z = -4.139$, $p < 0.01$).

DISCUSSION

This study aimed to investigate how auditory and somatosensory information is integrated in speech processing by school-aged children and adults, by testing three hypotheses.

As hypothesized, the overall perceptual categorization of the auditory stimuli was affected by the addition of somatosensory manipulations. The results for psychometric functions and categorical judgments revealed that auditory stimuli perceived simultaneously with skin stretch manipulations were labeled differently than when they were perceived on their own. Sounds were more perceived as /e/ when they were accompanied by the proprioceptive modification.

The second hypothesis that auditory and somatosensory integration would be greater in adults than in children was also confirmed. As shown in **Figures 2**, **3**, orofacial manipulation affected the position of the 50% crossover boundary of both groups; when backward skin stretches were perceived simultaneously with the auditory stimulus, it increased its probability of being identified as an /e/. This impact of skin stretch manipulation on the value corresponding to the 50th percentile was also reported in Ito et al.'s (2009) experiment. However, bimodal presentation of auditory and somatosensory inputs affected the steepness of the slope in adults only. **Figure 2** also shows that adult participants were more likely to label /e/-like stimuli as /e/ in the bimodal condition. Since negligible changes were observed for /ø/-like stimuli, it led to a more categorical boundary between the two acoustic vocalic targets. This difference in the integration patterns between children and adults suggests that linkage of specific somatosensory inputs with a corresponding speech sound evolves with age.

The third hypothesis that MSI would be stronger when auditory and somatosensory information was congruent was confirmed in adults but not in children. Only adults' perception was facilitated when both sensory information was consistent.

In children, a decrease in the correct identification rate resulted from the bimodal presentation when auditory and proprioceptive inputs were compatible. Moreover, while adults seemed to not be affected by the /e/-like skin stretches when auditory stimuli were alongside the prototypical /ø/ vocalic sound (see **Figure 2**), children's categorization was influenced even when sensory channels were clearly contrasting, as if the bimodal presentation of vocalic targets blurred the children categorization abilities. Moreover, thought somatosensory information mostly affected specific stimuli in adult, it's effect in children was further distributed along the auditory continuum. These last observations support our second hypothesis that MSI is strongly defined in adults.

As many have suggested, MSI continues to develop during childhood (e.g., Ross et al., 2011; Dionne-Dostie et al., 2015). The fact that young children are influenced by somatosensory inputs in a different manner then adults could, therefore, be due to their underdeveloped MSI abilities. Related findings have been reported for audiovisual integration (McGurk and MacDonald, 1976; Massaro, 1984; Desjardins et al., 1997). It has also been demonstrated that the influence of visual articulators in audition is weaker in school-aged children than in adults.

In agreement with the concept that MSI continues to develop during childhood, the differences observed between the two groups of perceivers could also be explained by the fact that different sensory systems develop at different rates and in different ways. In that sense, it has also been found that school-aged children were not only less likely to perceive a perceptual illusion resulting from incongruent auditory and visual inputs, but they also had poorer results in the identification of unimodal visual targets (Massaro, 1984).

Studies of the development of somatosensory abilities also support this concept. As established earlier, oral sensory acuity continues to mature until adolescence (McDonald and Aungst, 1967; Dette and Linke, 1982; Holst-Wolf et al., 2016). The young participants who were 5–6 years of age in the current study may have had underdeveloped proprioceptive systems, which may have caused their less clearly defined categorization of bimodal presentations.

It is generally accepted that auditory discrimination is poorer and more variable in children than in adults (Buss et al., 2009; MacPherson and Akeroyd, 2014), and children's lower psychometric scores are often related to poorer attention (Moore et al., 2008).

MSI requires sustained attention, and researchers have suggested that poor psychometric scores in children might be related to an attentional bias between the recruited senses in children vs. adults (Spence and McDonald, 2004; Alsius et al., 2005; Barutchu et al., 2009). For example, Barutchu et al. (2009) observed a decline in multisensory facilitation when auditory inputs were presented with a reduced signal-to-noise ratio. They suggested that the increased level of difficulty in performing the audiovisual detection task under high noise condition may be responsible for the degraded integrative processes.

If this attention bias might explain some of the between-group performance differences found when /e/-like somatosensory inputs were presented with /ø/-like auditory inputs (high level of difficulty), it would not justify differences between children and adults when the auditory and somatosensory channels agreed. The children showed decreased multisensory ability when both sensory inputs were compatible. Since difficulty level was reduced when multiple sensory sources were compatible, we should only have observed confusion in the children's categorization when auditory and somatosensory information was incongruent. According to the intersensory redundancy hypothesis, MSI should be improved when information from multiple sources is redundant. Indeed, Bahrick and Lickliter (2000) suggested that concordance of multiple signals would guide attention and even help learning (Barutchu et al., 2010). In the current study, this multisensory facilitation was only found in the adult participants.

This latter observation and the fact that no significant differences in variability were found across experimental conditions make it difficult to link the dissimilar patterns of MSI found between the two groups to an attention bias in children. However, finding a greater variability in MSI in children in both conditions, combined with their distinct psychometric and categorical scores provides support for the concept that perceptual systems in school-aged children are not yet fully shaped, which prevents them from attaining adult-like categorization scores.

As speech processing is multisensory and 5- to 6-year-olds have already experienced it, it is not surprising that some differences, even typical MSI ones, were found between the two experimental conditions in children. Since even very young children recognize that various speech sensory feedback can be compatible—or not (Patterson and Werker, 2003; Yeung and Werker, 2013; Bruderer et al., 2015; Werker, 2018), the different behavioral patterns observed in this study suggest that some form of multimodal processing exists in school-aged children, but complete maturation of the sensory systems is needed to achieve adult-like MSI.

CONCLUSION

When somatosensory input was added to auditory stimuli, it affected the categorization of stimuli at the edge of the categorical boundary for both children and adults. However, while the oral skin stretch manipulation had a defining effect on phonemic categories in adults, it seemed to have a blurring effect in children, particularly on the prototypical auditory stimuli. Overall, our results suggest that since adults have fully developed sensory channels and more experiences in MSI, they have stronger auditory and somatosensory integration than children.

Although longitudinal observations are not possible, two supplementary experiments in these participants has been conducted to further investigate how MSI takes place in speech processing in school-aged children and adults. These focus on the role of visual and auditory feedback.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Comité institutionnel d'éthique de la recherche avec des êtres humaines (CIERH) de l'Université du Québec à Montréal [UQAM; Institutional review board of research ethics with humans of the Université of Québec in Montréal (UQAM)]" with written informed consent from all subjects or their parent (for minor). All subjects or their parent gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Comité institutionnel d'éthique de la recherche avec des êtres humaines (CIERH) de l'Université du Québec à Montréal [UQAM; Institutional review board of research ethics with humans of the Université of Québec in Montréal (UQAM)]".

AUTHOR CONTRIBUTIONS

PT-F, TI and LM contributed to the conception and design of the study. PT-F collected data, organized the database and performed the statistical analysis (all under LM's guidance). PT-F wrote the first draft of the manuscript. PT-F and LM were involved in subsequent drafts of the manuscript. PT-F, TI and LM contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was funded by the Social Sciences and Humanities Research Council of Canada (both Canadian Graduate Scholarships—a Doctoral program and an Insight grant) and the Natural Sciences and Engineering Research Council of Canada (a Discovery grant).

ACKNOWLEDGMENTS

We are especially grateful to David Ostry for his immense generosity toward this project and to Marlene Busko for copyediting the article. Special thoughts for Bilal Alchalabi and Camille Vidou for their precious help.

REFERENCES

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843. doi: 10.1016/j.cub.2005.03.046

Arabin, B. (2002). Music during pregnancy. *Ultrasound Obstet. Gynecol* 20, 425–430. doi: 10.1046/j.1469-0705.2002.00844.x

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Bahrick, L. E., and Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev. Psychol.* 36, 190–201. doi: 10.1037//0012-1649.36.2.190

Bahrick, L. E., and Lickliter, R. (2012). "The role of intersensory redundancy in early perceptual, cognitive and social development," in *Multisensory Development*, eds A. Bremner, D. J. Lewkowicz and C. Spence (Oxford: Oxford University Press), 183–205.

Barutchu, A., Crewther, D. P., and Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Dev. Sci.* 12, 464–473. doi: 10.1111/j.1467-7687.2008.00782.x

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., and Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *J. Exp. Child Psychol.* 105, 38–50. doi: 10.1016/j.jecp.2009.08.005

Birch, H. G., and Lefford, A. (1963). Intersensory development in children. *Monogr. Soc. Res. Child Dev.* 28, 1–47. doi: 10.2307/1165681

Boë, L.-J., Granat, J., Badin, P., Autesserre, D., Pochic, D., Zga, N., et al. (2007). "Skull and vocal tract growth from newborn to adult," in *7th International Seminar on Speech Production, ISSP7* (Ubatuba, Brazil), 533–536. Available online at: https://hal.archives-ouvertes.fr/hal-00167610. Accessed March 15, 2019.

Boë, L.-J., Ménard, L., Serkhane, J., Birkholz, P., Kröger, B., Badin, P., et al. (2008). La croissance de l'instrument vocal: contrôle, modélisation, potentialités acoustiques et conséquences perceptives. *Revue Française de Linguistique Appliquée* 13, 59–80. Available online at: https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2008-2-page-59.htm. Accessed March 15, 2019.

Bower, T. G., Broughton, J. M., and Moore, M. K. (1970). The coordination of visual and tactual input in infants. *Percept. Psychophys.* 8, 51–53. doi: 10.3758/bf03208933

Bruderer, A. G., Danielson, D. K., Kandhadai, P., and Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proc. Natl. Acad. Sci. U S A* 112, 13531–13536. doi: 10.1073/pnas.1508631112

Burr, D., and Gori, M. (2011). "Multisensory integration develops late in humans," in *The Neural Bases of Multisensory Processes*, eds M. M. Murray and M. T. Wallace (Boca Raton, FL: CRC Press), 810.

Buss, E., Hall, J. W., and Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: slope differences in school-aged children and adults. *J. Acoust. Soc. Am.* 125, 1050–1058. doi: 10.1121/1.3050273

Demany, L., McKenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature* 266, 718–719. doi: 10.1038/266718a0

DePaolis, R. A., Vihman, M. M., and Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behav. Dev.* 34, 590–601. doi: 10.1016/j.infbeh.2011.06.005

Desjardins, R. N., Rogers, J., and Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *J. Exp. Child Psychol.* 66, 85–110. doi: 10.1006/jecp.1997.2379

Dette, M., and Linke, P. G. (1982). The development of oral and manual stereognosis in children from 3 to 10 years old (Die Entwicklung der oralen und manuellen Stereognose bei Kindem im Altervon 3 bis 10 Jahren). *Stomatologie* 32, 269–274.

Dionne-Dostie, E., Paquette, N., Lassonde, M., and Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sci.* 5, 32–57. doi: 10.3390/brainsci5010032

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. Available online at: http://home.fau.edu/lewkowic/web/Eimas infant speech discrim Science 1971.pdf. Accessed February 1, 2019.

Feng, Y., Gracco, V. L., and Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *J. Neurophysiol.* 106, 667–679. doi: 10.1152/jn.00638.2010

Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522. doi: 10.1121/1.427148

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1152/jn.2002.88.1.540

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572

Gisel, E. G., and Schwob, H. (1988). Oral form discrimination in normal 5-to 8-year-old children: an adjunct to an eating assessment. *Occup. Ther. J. Res.* 8, 195–209. doi: 10.1177/153944928800800403

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036

Guenther, F. H., and Perkell, J. S. (2004). "A neural model of speech production and its application to studies of the role of auditory feedback in speech," in *Speech Motor Control in Normal and Disordered Speech*, eds R. Maassen, R. D. Kent, H. Peters, P. H. van Lieshout and W. Hulstijn (Oxford: Oxford University Press), 29–49.

Guenther, F. H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. *J. Neurolinguistics* 25, 408–422. doi: 10.1016/j.jneuroling.2009.08.006

Guest, S., Catmur, C., Lloyd, D., and Spence, C. (2002). Audiotactile interactions in roughness perception. *Exp. Brain Res.* 146, 161–171. doi: 10.1007/s00221-002-1164-z

Hatwell, Y. (1987). Motor and cognitive functions of the hand in infancy and childhood. *Int. J. Behav. Dev.* 10, 509–526. doi: 10.1177/016502548701000409

Hecht, D., and Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Exp. Brain Res.* 193, 307–314. doi: 10.1007/s00221-008-1626-z

Hillock, A. R., Powers, A. R., and Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia* 49, 461–467. doi: 10.1016/j.neuropsychologia.2010.11.041

Holst-Wolf, J. M., Yeh, I.-L., and Konczak, J. (2016). Development of proprioceptive acuity in typically developing children: normative data on forearm position sense. *Front. Hum. Neurosci* 10:436. doi: 10.3389/fnhum.2016.00436

Horlyck, S., Reid, A., and Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: maturation versus experience. *Sci. Stud. Read.* 16, 218–239. doi: 10.1080/10888438.2010.546460

Hötting, K., and Röder, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychol. Sci.* 15, 60–64. doi: 10.1111/j.0963-7214.2004.01501010.x

Ito, T., and Gomi, H. (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *Neuroreport* 18, 907–910. doi: 10.1097/WNR.0b013e32810f2dfb

Ito, T., and Ostry, D. J. (2010). Somatosensory contribution to motor learning due to facial skin deformation. *J. Neurophysiol.* 104, 1230–1238. doi: 10.1152/jn.00199.2010

Ito, T., Gracco, V. L., and Ostry, D. J. (2014). Temporal factors affecting somatosensory-auditory interactions in speech processing. *Front. Psychol.* 5:1198. doi: 10.3389/fpsyg.2014.01198

Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U S A* 106, 1245–1248. doi: 10.1073/pnas.0810063106

Johansson, R. S., Trulsson, M., Olsson, K. A., and And Abbs, J. H. (1988). Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Exp. Brain Res.* 72, 209–214. doi: 10.1007/bf00248519

Jousmäki, V., and Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Curr. Biol.* 8:R190. doi: 10.1016/s0960-9822(98)70120-4

Katseff, S., Houde, J. F., and Johnson, K. (2012). Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Lang. Speech* 55, 295–308. doi: 10.1177/0023830911417802

Krakauer, J. W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., and Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS Biol.* 4:e316. doi: 10.1371/journal.pbio.0040316

Krueger, L. E. (1970). David Katz's der aufbau der tastwelt (the world of touch): a synopsis. *Percept. Psychophys.* 7, 337–341. doi: 10.3758/BF03208659

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Kuhl, P., and Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science* 218, 1138–1141. doi: 10.1126/science.7146899

Kumin, L. B., Saltysiak, E. B., Bell, K., Forget, K., Goodman, M. S., Goytisolo, M., et al. (1984). Relationships of oral stereognostic ability to age and sex of children. *Percept. Mot. Skills* 59, 123–126. doi: 10.2466/pms.1984.59.1.123

Lametti, D. R., Nasir, S. M., and Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J. Neurosci.* 32, 9351–9358. doi: 10.1523/JNEUROSCI.0404-12.2012

Lecanuet, J.-P., Granier-Deferre, C., and Busnel, M.-C. (1995). "Human fetal auditory perception," in *Fetal Development: A Psychobiological Perspective*, eds J.-P. Lecanuet, W. P. Fifer, N. A. Krasnegor and W. P. Smotherman (Hillsdale, NJ: Lawrence Erlbaum Associates), 512.

Macaluso, E., Frith, C. D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science.* 289, 1206–1208. doi: 10.1126/science.289.5482.1206

MacDonald, E. N., Goldberg, R., and Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *J. Acoust. Soc. Am.* 127, 1059–1068. doi: 10.1121/1.3278606

MacPherson, A., and Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear.* 18:2331216514537722. doi: 10.1177/2331216514537722

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev.* 55, 1777–1788.

Massaro, D. W. (1999). Speech reading: illusion or window into pattern recognition. *Trends Cogn. Sci.* 3, 310–317. doi: 10.1016/S1364-6613(99)01360-1

McDonald, E. T., and Aungst, L. F. (1967). "Studies in oral sensorimotor function," in *Symposium on Oral Sensation and Perception*, ed. J. F. Bosma (Springfield, IL: Charles C. Thomas), 202–220.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

McGurk, H., and Power, R. P. (1980). Intermodal coordination in young children: vision and touch. *Dev. Psychol.* 16, 679–680. doi: 10.1037/0012-1649.16.6.679

Ménard, L., and Boe, L. (2004). L'émergence du système phonologique chez l'enfant : l'apport de la modélisation articulatoire. *Cana. J. Linguist.* 49, 155–174. doi: 10.1017/S0008413100000785

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., and Molholm, S. (2013). Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage* 79, 19–29. doi: 10.1016/j.neuroimage.2013.04.060

Misceo, G. F., Hershberger, W. A., and Mancini, R. L. (1999). Haptic estimates of discordant visual-haptic size vary developmentally. *Percept. Psychophys.* 61, 608–614. doi: 10.3758/bf03205533

Mishra, J., Martinez, A., Sejnowski, T. J., and Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J. Neurosci.* 27, 4120–4131. doi: 10.1523/JNEUROSCI.4912-06.2007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128. doi: 10.1016/s0926-6410(02)00066-6

Moore, D. R., Ferguson, M. A., Halliday, L. F., and Riley, A. (2008). Frequency discrimination in children: perception, learning and attention. *Hear. Res.* 238, 147–154. doi: 10.1016/j.heares.2007.11.013

Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Curr. Biol.* 18, 689–693. doi: 10.1016/j.cub.2008.04.021

Nasir, S. M., and Ostry, D. J. (2006). Somatosensory precision in speech production. *Curr. Biol.* 16, 1918–1923. doi: 10.1016/j.cub.2006.07.069

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., and Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Dev. Sci.* 9, 454–464. doi: 10.1111/j.1467-7687.2006.00512.x

Ogane, R., Schwartz, J.-L., and Ito, T. (2017). "Somatosensory information affects word segmentation and perception of lexical information," in *(Poster) Presented at the 2017 Society for the Neurobiology of Language conference in Baltimore, MD.* Available online at: https://hal.archives-ouvertes.fr/hal-01658527. Accessed March 05, 2019.

Patterson, M. L., and Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* 6, 191–196. doi: 10.1111/1467-7687.00271

Perrier, P. (1995). Control and representations in speech production. *ZAS Papers Lingustics* 40, 109–132. Available online at: https://hal.archives-ouvertes.fr/hal-00430387/document. Accessed March 05, 2019.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., and Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U S A* 106, 10598–10602. doi: 10.1073/pnas.0904134106

Proske, U., and Gandevia, S. C. (2009). The kinaesthetic senses. *J. Physiol.* 587, 4139–4146. doi: 10.1113/jphysiol.2009.175372

Purcell, D. W., and Munhall, K. G. (2006). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J. Acoust. Soc. Am.* 120, 966–977. doi: 10.1121/1.2217714

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., et al. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur. J. Neurosci* 31, 1772–1782. doi: 10.1111/j.1460-9568.2010.07213.x

Rentschler, I., Jüttner, M., Osman, E., Müller, A., and Caelli, T. (2004). Development of configural 3D object recognition. *Behav. Brain Res.* 149, 107–111. doi: 10.1016/s0166-4328(03)00194-3

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T., and Escudier, P. (1998). Complementarity and synergy in bimodal speech: auditory, visual and audio-visual identification of French oral vowels in noise. *J. Acoust. Soc. Am.* 103, 3677–3689. doi: 10.1121/1.423069

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-amour, D., and Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *Eur. J. Neurosci.* 33, 2329–2337. doi: 10.1111/j.1460-9568.2011.07685.x

Schürmann, M., Caetano, G., Jousmäki, V., and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Am.* 115, 830–832. doi: 10.1121/1.1639909

Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399. doi: 10.1093/cercor/bhl147

Spence, C., and McDonald, J. (2004). "The cross-modal consequences of the exogenous spatial orienting of attention," in *The Handbook of Multisensory Processes*, eds G. A. Calvert, C. Spence and B. E. Stein (Cambridge, MA: MIT Press), 3–25.

Stein, B. E., London, N., Wilkinson, L. K., and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci.* 8, 497–506. doi: 10.1162/jocn.1996.8.6.497

Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. E., Stanford, T. R., and Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nat. Rev. Neurosci.* 15, 520–535. doi: 10.1038/nrn3742

Streri, A., and Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia* 42, 1365–1369. doi: 10.1016/j.neuropsychologia.2004.02.012

Tourville, J. A., and Guenther, F. H. (2011). The DIVA model: a neural theory of speech acquisition and production. *Lang. Cogn. Process.* 26, 952–981. doi: 10.1080/01690960903498424

Tremblay, S., Shiller, D. M., and Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature* 423, 866–869. doi: 10.1038/nature01710

Trudeau-Fisette, P., Tiede, M., and Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: acoustic and articulatory data. *PLoS One* 12:e0180300. doi: 10.1371/journal.pone.0180300

Turgeon, C. (2011). Mesure du développement de la capacité de discrimination auditive et visuelle chez des personnes malentendantes porteuses d'un implant cochléaire. Available online at: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/6091. Accessed February 01, 2019.

Vihman, M. M. (1996). *Phonological Development: The Origins of Language in the Child.* Oxford, England: Blackwell.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J. Acoust. Soc. Am.* 122, 2306–2319. doi: 10.1121/1.2773966

Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results. *Int. J. Pediatr. Otorhinolaryngol.* 49, 197–206. doi: 10.1016/s0165-5876(99)00208-6

Walker-Andrews, A. (1994). "Taxonomy for intermodal relations," in *The Development of Intersensory Perception: Comparative Perspectives*, eds D. J. Lewkowicz and R. Lickliter (Hillsdale, NJ: Lawrence Erlbaum Associates), 39–56.

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Appl. Psycholinguist* 39, 703–728. doi: 10.1017/s014271641800022x

Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* 24, 603–612. doi: 10.1177/0956797612458802

Yu, L., Rowland, B. A., and Stein, B. E. (2010). Initiating the development of multisensory integration by manipulating sensory experience,. *J. Neurosci.* 30, 4904–4913. doi: 10.1523/JNEUROSCI.5575-09.2010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 25 October 2019
doi: 10.3389/fnhum.2019.00380

Common Neural System for Sentence and Picture Comprehension Across Languages: A Chinese–Japanese Bilingual Study

Zhengfei Hu[1]*, Huixiang Yang*[1]*, Yuxiang Yang*[1]*, Shuhei Nishida*[1]*, Carol Madden-Lombardi*[2]*, Jocelyne Ventre-Dominey*[2]*, Peter Ford Dominey*[2] *and Kenji Ogawa*[1]*

[1] *Department of Psychology, Hokkaido University, Sapporo, Japan,* [2] *INSERM - U1093 Cognition, Action, and Sensorimotor Plasticity, Dijon, France*

While common semantic representations for individual words across languages have been identified, a common meaning system at sentence-level has not been determined. In this study, fMRI was used to investigate whether an across-language sentence comprehension system exists. Chinese–Japanese bilingual participants ($n = 32$) were asked to determine whether two consecutive stimuli were related (coherent) or not (incoherent) to the same event. Stimuli were displayed with three different modalities (Chinese written sentences, Japanese written sentences, and pictures). The behavioral results showed no significant difference in accuracy and response times among the three modalities. Multi-voxel pattern analysis (MVPA) of fMRI data was used to classify the semantic relationship (coherent or incoherent) across the stimulus modalities. The classifier was first trained to determine coherency within Chinese sentences and then tested with Japanese sentences, and vice versa. A whole-brain searchlight analysis revealed significant above-chance classification accuracy across Chinese and Japanese sentences in the supramarginal gyrus (BA 40), extending into the angular gyrus (BA 39) as well as the opercular (BA 44) and triangular (BA 45) parts of the inferior frontal gyrus in the left hemisphere (cluster-level FWE corrected $p < 0.05$). Significant above-chance classification accuracy was also found across Japanese sentences and pictures in the supramarginal (BA 40) and angular gyrus (BA 39). These results indicate that a common meaning system for sentence processing across languages and modalities exists, and it involves the left inferior parietal gyrus.

Keywords: semantic processing, sentence comprehension, bilingualism, fMRI, MVPA

Edited by:
Hidehiko Okamoto,
International University of Health and Welfare (IUHW), Japan

Reviewed by:
Andrea Leo,
IMT School for Advanced Studies Lucca, Italy
Bernadette Maria Jansma,
Maastricht University, Netherlands

Correspondence:
Kenji Ogawa
ogawa@let.hokudai.ac.jp

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *25 June 2019*
Accepted: *11 October 2019*
Published: *25 October 2019*

Citation:
Hu Z, Yang H, Yang Y, Nishida S, Madden-Lombardi C, Ventre-Dominey J, Dominey PF and Ogawa K (2019) Common Neural System for Sentence and Picture Comprehension Across Languages: A Chinese–Japanese Bilingual Study. Front. Hum. Neurosci. 13:380. doi: 10.3389/fnhum.2019.00380

INTRODUCTION

Some of the languages in existence nowadays share similarities in phonological and/or orthographic properties, while others do not. However, all human beings are capable of acquiring another language besides their native language. Language is a symbolic representation of the knowledge of the world, the meaning which is also known as semantics in the domain of linguistics. It is possible to assume that the neurobiological infrastructure that is largely shared among humans is likely to be the neural system that underlines semantic processing (Binder et al., 2009; Hagoort, 2014).

The comprehension of semantics requires the automatic parallel processing of sound, word, and sentence patterns (Fromkin et al., 2014). Nevertheless, the semantic properties of words/sentences are readily distinguished from their structural properties (Binder et al., 2009). Also, the neural processing of different language structures is distinguishable (e.g., Tan et al., 2005; Grodzinsky and Friederici, 2006; Buchweitz et al., 2009), even as the brain regions overlap to some degree (e.g., Keller et al., 2001; Hickok and Poeppel, 2004).

Binder et al. (2009) reviewed neuroimaging studies (i.e., fMRI and PET studies) to identify brain regions that contribute to the semantic component of words and found that the left posterior parietal lobe, the lateral temporal cortex, and the inferior frontal gyrus demonstrated a high likelihood of activation across studies. Recent fMRI studies applied multi-voxel pattern analysis (MVPA) to investigate neural representations associated with semantics by analyzing patterns of neural activation (Mitchell et al., 2008; Shinkareva et al., 2011). This approach also enabled a comparison through which to investigate the common neural representations across people and languages (Zinszer et al., 2016; Yang et al., 2017a), especially in bilingual semantic processing. For example, the common neural representation of equivalent Portuguese and English nouns was found to be situated in the left post-central and supramarginal gyri (SMG), the left inferior and superior parietal lobes (I/SPL), the left inferior frontal gyrus (IFG), and the posterior superior temporal lobe (Buchweitz et al., 2012). Meanwhile Correia et al. (2014) argued that the shared representation across Dutch and English was located in the left anterior temporal lobe (ATL), the left angular gyrus (AG), the posterior bank of the left postcentral gyrus, the right posterior superior temporal sulcus/gyrus (STS/STG), the right anterior insula, the medial part of right ATL, and the bilateral occipital cortices. Van de Putte et al. (2017) investigated the common neural representation across French and Dutch and proved that the shared semantic representations are located in the bilateral occipito-temporal cortex and in the inferior and the middle temporal gyrus (ITG/MTG). Overall, the previous neuroimaging studies investigating the common neural representation of the semantic processing of words have yielded a consistent result, regardless of whether the participants were asked to read (Buchweitz et al., 2012), listen (Correia et al., 2014), or speak (Van de Putte et al., 2017) the words. Those studies suggested that a common neural representation might comprise a number of brain regions, including the left inferior parietal lobe (AG and portions of SMG) and the superior/middle temporal lobe. However, inconsistent results still existed, and this inconsistency might be resolved via using more natural processing, i.e., sentence processing.

In real life we communicate in written or spoken sentences formed of words that are arranged according to complicated syntactic rules (Ingram, 2007). Accordingly, the meanings conveyed by sentences transcend the individual words. However, the neural system underlying the semantic processing of sentences is still controversial. Price (2010) reviewed studies that have investigated the brain regions involved in the semantic processing of spoken sentences and argued that the neural system at the sentence level was situated in the anterior and posterior parts of the left middle temporal gyrus, the bilateral anterior temporal poles, the left AG, and the posterior cingulate/precuneus. These areas were associated with the semantic processing of words as reported by Binder et al. (2009). Interestingly, Jouen et al. (2015) identified a common neural system from processing whole sentences and images that describe human events that also includes the left AG. To our knowledge, however, only one MVPA study has argued the existence of commonalities in the neural system of bilinguals in the semantic processing of sentences across languages: Yang et al. (2017b) mapped the semantic properties of English words and their neural representations and subsequently developed a predictive model containing the neural system of sentences that were composed from these words. Although they demonstrated the above chance accuracy of predicting the activation pattern of Portuguese sentences from equivalent English sentences, the Portuguese–English bilinguals were presented with only their native Portuguese sentences. The direct prediction of simultaneous semantic processing between the two languages known by the bilinguals was not conducted. Also, as Yang et al. (2017b) have suggested, all the across-language neural decoding and prediction studies used stimuli that only encompassed a small semantic space. In other words, they used a limited number of concrete nouns to represent dwellings, tools, animals, or other objects. It is hence unknown whether the neural system of the vast semantic spaces across languages can be similarly predicted.

In order to achieve the understanding of the semantics of sentences precisely, the syntactic processing is necessary for dealing with fitnesses of different arguments of words and phrases (Bookheimer, 2002; Humphries et al., 2007). This syntactic processing is considered to be subserved by the pars opercularis (BA 44), a subpart of the left inferior frontal gyrus (IFG) (Newman et al., 2010; Friederici, 2011; Makuuchi and Friederici, 2013). Further, during the sentence comprehension, the syntactic information needs to be integrated with the semantic information (Marslen-Wilson and Tyler, 1980) that is subserved by the pars trianularis (BA 45) which is another subpart of the left IFG (Friederici, 2011). Thus, the integration processing of the syntactic and semantic information is assumed to be supported in the left IFG including BA 44 and BA 45 (Hagoort, 2005, 2014). However, some findings suggested that the region which supports the processing of the syntactic and semantic integration is located in the posterior temporal cortex (Friederici, 2011, 2012). Despite these controversies, we hypothesized that the common neural system of the semantic processing of sentences might comprise regions associated with the syntactic processing, which are located in the left IFG.

This study aimed to investigate the common neural system of the semantic processing of sentences across languages. Bilingual participants were asked to read both the Chinese and the equivalent Japanese sentences and to understand them. The cross-language classification was implemented. This application comprised training the support vector machine (SVM) classifier

(Cortes and Vapnik, 1995) with sentences in one language and testing it with sentences in the other language and vice versa. Our analysis involved the training and testing of the SVM with sentences in one language. In addition, participants were presented with pictures that depicted the same kinds of human events as in the sentences. The participants thus performed the semantic processing of three different modalities: Chinese written sentences, Japanese written sentences, and pictures.

MATERIALS AND METHODS

Participants

Thirty-two right-handed speakers of Chinese as their first language participated in the study. Behavioral results showed that the accuracy of three participants on either task condition was lower than the chance level (50%). Thus, these three participants were removed from both the behavioral and the fMRI analyses. The remaining 29 participants (6 males; mean age = 27.93, *SD* = 3.65) with normal or corrected to normal vision reported that they did not suffer from any neurological or psychiatric disorder. Each participant signed the informed consent format approved by the Ethical Committee of the Graduate School of Letters at Hokkaido University.

All the participants were late sequential bilinguals who began learning Japanese at the average age of 18.03 (*SD* = 3.26). They had been learning Japanese for 9.71 years (*SD* = 4.23) on average, and had been living in Japan for an average of 4.85 years (*SD* = 2.59). Except for one participant who had studied and lived in Japan for more than 10 years, all the other participants had passed the highest level of Japanese Language Proficiency Test. Twenty-two of the participants were enrolled in a graduate-level course, and one was registered in an undergraduate program at Hokkaido University. Seven of the participants were employed in occupations: of these, four were master's degree holders, two had earned doctoral degree, and one passed the requirements for the bachelor's degree.

The participants filled out a language self-rating questionnaire to help researchers ascertain their Japanese language proficiency. The questionnaire asked participants to award self-rating points on a scale of 1.0 (poor) to 7.0 (excellent). The questionnaire contained five questions each on listening, speaking, reading, and writing skills which were collected from the JLPT Can-do Self-Evaluation List created by the Japan Foundation and the Japan Educational Exchanges and Services[1]. Despite being late bilinguals, the participants rated themselves as being highly proficient in Japanese (*M* = 6.24, *SD* = 0.58).

Stimuli

The stimuli comprised 48 pairs of pictures, and Chinese and Japanese written sentences totaling 144 pairs. The pictures (adapted from Jouen et al., 2015) depicted events representing one or two persons (no negative emotional valence) performing a common daily activity (e.g., playing the piano, cooking, reading a book to a child, etc.) and were collected from the Getty photo database[2]. A pair of pictures either symbolized a sequence of coherent events (for example, the first picture showed a girl throwing a piece of rock and the second picture portrayed the girl playing hopscotch as shown in **Figure 1**) or incoherent events (e.g., the first displayed a girl and a woman mounting wallpaper, and the second portrayed the same two characters jumping up and down on a bed), and the paired pictures represented only either coherent or incoherent events. Both the Chinese and Japanese sentences were generated on the basis of the pictures: the sentences described the activities being performed by the people in the pictures. The Chinese sentences were first generated and subsequently translated into equivalent Japanese sentences. The validity of the Japanese translation was confirmed by consulting with a native Japanese speaking expert. As a result, each event pair conveying the same meanings was represented by three different modalities: the picture, the Chinese sentence, and the Japanese sentence.

Three stimuli sets (A, B, and C) were generated, each comprising 16 picture pairs, 16 Chinese sentence pairs, and 16 Japanese sentence pairs. To avoid the participants seeing a picture pair and receiving its corresponding Chinese or Japanese sentence pair in a single session, stimuli pairs were crossed. For example, the 16 picture pairs were classified into set A, and their corresponding 16 Chinese and Japanese sentence pairs were then respectively classified into sets B and C.

The stimuli were balanced for coherency (coherent or incoherent) and the number of individuals performing the activity (one or two), and both were counterbalanced across the three stimuli sets. The Chinese sentences obeyed the subject–verb–object order and the Japanese sentences obeyed the subject–object–verb order according to the grammar rules of the respective languages. The Chinese sentences had a mean length of 10.08 words (*SD* = 2.62) and the Japanese sentences had an average of 16.33 words (*SD* = 3.38).

Stimuli Evaluation

To assess the degree to which the sentences matched with the pictures, 20 pilot participants (10 Chinese natives and 10 Japanese natives) who did not participate in the fMRI experiment were recruited. These participants were presented with sentences in their respective native language and were asked to rate how appropriately the sentences were able to describe the activities being performed by the persons in the pictures on a scale of 1.0 (very poorly) to 7.0 (very well). The mean rating score was 6.51 (*SD* = 0.40) for the Chinese sentences and 6.69 (*SD* = 0.30) for the Japanese sentences. Both the Chinese and the Japanese sentences were thus rated as being excellently able to describe the activities in the pictures. In addition, the variety of activities in the pictures ensured that the Japanese used in the sentences would incorporate a vast semantic space. Insufficient knowledge of the used Japanese sentences would probably affect the semantic processing; therefore, the frequency of the use of the words of the Japanese sentences was evaluated. All the Japanese sentences were segmented into words (totally 179 words), and the incidence

---

[1] http://www.jlpt.jp/about/candolist.html

[2] https://www.gettyimages.fr/

FIGURE 1 | Experimental paradigm.

of the use of the particular words was investigated using the long-unit-word and the short-unit-word aspect of The Balanced Corpus of Contemporary Written Japanese (BCCWJ) (National Institute for Japanese Language and Linguistics, 2018). Except for the newly coined word "selfie," the frequencies of the words used for the sentences ranked between extremely low and high levels of usage, extending from 0.16 to 48383.91 per million words.

Procedures

Before the fMRI scanning session, participants completed the vocabulary checking list that was generated according to word frequency. The list contained 30 words collected in the order of decreasing word frequency. The participants were asked to remember the meanings of the words that they did not know to avoid interference with sentence comprehension caused by not understanding the words that were used.

During the scanning, participants underwent three sessions (i.e., stimuli sets A, B, and C), and the order of the sessions was counterbalanced across participants. Each session included 48 trials, and all the trials were randomly presented. During a trial, a yellow fixation cross appeared on a black background for 1 s, then the first stimulus of the event pair appeared on a black background for 3 s, followed by another yellow fixation cross for 1 s before the second stimulus of the event pair was presented for 3 s. At the end of this sequence, an evaluation screen with a white fixation cross containing options on a black background was presented for 2 s plus jitter time (0, 2, 4, or 6 s) before the next trial (**Figure 1**). In the trials in which event pairs were presented either in Chinese or Japanese, participants were instructed to silently and consistently read the sentences until they disappeared. In the trials in which the event pairs were presented in pictures, participants were instructed to continue thinking about the activities that the persons in the pictures were performing. In each trial, participants were also asked to judge whether or not the event pair was coherent by pressing a response button with their right index or middle finger as soon as they could after the evaluation screen appeared. The evaluation screen did not disappear after they had pressed this button. The judging options were arranged in accordance with the participants' fingers: the alternative presented in the left bottom corner corresponded to the index finger and the selection appearing in the right bottom corner corresponded to the middle finger. The locations of the options were counterbalanced across the stimuli.

fMRI Acquisition

Functional and structural image acquisition was performed on a Siemens Prisma 3.0 T scanner using a 64-channel head coil at the Research and Education Center for Brain Science of Hokkaido University. The whole brain functional images were collected using a T2*-weighted gradient EPI sequence (TR = 2000 ms, TE = 30 ms, voxel size = 2 × 2 × 3.5 mm, and a 90° flip angle). A session consisted of 318 volumes. The high-resolution structural images covering the whole brain were acquired after the functional image acquisition using a T1 MPRAGE sequence (TR = 2300 ms, TE = 2.41 ms, TI = 900 ms, FOV = 256 × 256 mm, and an 8° flip angle).

fMRI Data Preprocessing

fMRI data processing and analysis were performed with SPM 12 (Wellcome Department of Cognitive Neurology, London, United Kingdom) in the Matlab environment. The first three scans of all the sessions were removed from the analysis to minimize T1 artifacts. The functional images were corrected for slice timing and were spatially realigned to normalize to the Montreal Neurological Institute (MNI) space without changing the voxel size. Spatial smoothing was applied using a Gaussian kernel of 6 × 6 × 6 mm full width at half-maximum for univariate analysis. To prevent the possibility of less predictive individual voxels, spatially normalized but unsmoothed images were used to perform MVPA.

fMRI Data Analysis: Univariate Analysis

To reveal neural regions generally involved in the semantic processing of Chinese sentences, Japanese sentences, and pictures, the general linear model (GLM) (Friston et al., 1995) was used to obtain contrasts between each modality and the baseline. The three different modalities were modeled as three separate regressors and were convolved with a canonical hemodynamic response function for each participant. The group analysis was then processed through a second-level random effects model by using a one-sample t-test in a group analysis of all the participants. The activated regions were extracted with a cluster-level ($k \geq 15$) threshold of $p < 0.05$ corrected for family-wise error (FWE).

fMRI Data Analysis: Searchlight MVPA

A searchlight method (Kriegeskorte et al., 2006) with a linear SVM classifier, as implemented by LIBSVM (Chang and Lin, 2011), was performed to investigate the common semantic neural system across languages and modalities in the processing of sentence comprehension. A spherical searchlight with a radius of 9 mm was used to reveal multiple patterns that carried featured neural representations of the sentence semantics.

The classifier for the classification analysis was trained to discriminate between the neural patterns associated with the coherent and incoherent events. Two types of classification analysis were performed. First, a within-language/modality classification was accomplished with (1) only Chinese sentence trials, (2) only Japanese sentence trials, and (3) only picture trials. In all the within-language/modality classifications, a leave-one-out procedure was used: the classifier was trained on the data from any two of the three sessions and was tested on the data from the one session left. The classification was repeated thrice by interchanging the training and testing data. The accuracies were averaged across the three iterations.

The second cross-classification was accomplished across languages/modalities classification with (4) Chinese vs. Japanese sentences, (5) Chinese sentence vs. picture, and (6) Japanese sentence vs. picture. The classifier was trained with data from one language/modality condition belonging to all the three sessions. It was tested on the respective data obtained from the remaining language/modality condition belonging to all the three sessions. Each classification was repeated twice in a manner ensuring that all of the specific language/modality data were used once for the test. The resulting accuracies were averaged across classification directions.

To construct accuracy group maps for across-language/modality, the accuracies were averaged across all participants and contrasted with the average accuracy of the coherency (accuracy at chance = 50%) using a one-sample t-test to reveal the cluster level ($k \geq 10$) significant classification of sentence semantics across languages/modalities ($p < 0.05$, FWE-corrected). Group maps were also produced for the within-language/modality ($p < 0.05$, FWE-corrected, $k \geq 10$).

To reveal a more robust result for the searchlight MVPA analysis, statistical maps were corrected using threshold free cluster enhancement (TFCE; Smith and Nichols, 2009) as implemented in a free MatlabTFCE package[3] which combined a maximal permuted statistic correction technique (Nichols and Holmes, 2001). Ten thousand permutations and a one-tailed corrected cluster threshold of p = 0.05 were used (Wurm and Caramazza, 2019).

fMRI Data Analysis: Region of Interest (ROI) Analysis

An ROI analysis was further performed to specifically investigate the effects of regions that are commonly activated during sentence processing. ROIs were selected based on Jouen et al. (2015) study, which investigated the common neural representations for the semantic processing of sentences in monolinguals by performing conjunction analysis between written sentences and pictures. Regions such as BA 22 (superior temporal gyrus), BA 39 (AG, inferior parietal lobe), and BA 45 (triangular part of inferior frontal gyrus) were selected on the basis of their reporting of these areas being involved in the semantic processing of sentences. Two further regions were selected: BA 40 (supramarginal gyrus), which is adjacent to BA 39, and BA 44 (the opercular part of the inferior frontal gyrus), which is adjacent to BA 45. These regions are considered to be part of the classical language region and are known as Wernicke's area and Broca's area, respectively (Catani et al., 2005; Hagoort, 2017). To assure that the sentences were processed as semantic stimuli rather than visual stimuli, BA 17 (primary visual cortex) was also

---

[3] https://github.com/markallenthornton/MatlabTFCE

selected, and it was proved to be involved in visual information processing as the control region.

RESULTS

Behavioral Results

During the fMRI scanning, the participants performed a coherence judging task. In all the conditions involving the Chinese sentence, the Japanese sentence and the picture, they evaluated the coherence with a high accuracy (Chinese sentence: $M = 92\%$, $SEM = 0.01$; Japanese sentence: $M = 92\%$, $SEM = 0.01$; picture: $M = 93\%$, $SEM = 0.01$; **Figure 2A**). No significant differences were found on the accuracy [$F(2,56) = 0.56$, $p = 0.58$, $\eta_p^2 = 0.02$].

A significant difference was found pertaining to the condition on the response time [$F(2,56) = 26.95$, $p < 0.001$, $\eta_p^2 = 0.49$]. Participants responded faster to the picture (M=0.94s, SEM = 0.03) than to the Chinese sentence (M=1.00s, SEM=0.04) and to the Japanese sentence (M=1.06s, SEM=0.04; **Figure 2B**). In a supplementary analysis, differences in the accuracy and the response time of the coherence judgment for all the conditions were analyzed (see **Supplementary Material**).

Univariate Analysis

We accomplished a voxel-based analysis of the whole brain activation to reveal activated neural regions for semantic processing of the Chinese sentence, the Japanese sentence, and the picture ($p < 0.05$, FWE-corrected, ke $\geq$ 15).

For the semantic processing of the Chinese sentence, a large predominantly left-hemisphere network was activated (**Figure 3** and **Table 1A**). These regions included clusters spreading from the precentral gyrus (BA 6) and the supplementary motor area (SMA; BA 6) to the opercular part of the inferior frontal gyrus (Oper-IFG; BA 44); from the lateral inferior occipital gyrus (IOG; BA 18) to the fusiform gyrus (BA 37); and from the middle temporal gyrus (MTG; BA 21/22) to the superior parietal gyrus (SPG; BA 7) in the left hemisphere. The clusters that spread from the fusiform gyrus (BA 37) to the lingual gyrus (BA 18) and calcarine cortex (BA 17), and areas which included the MTG (BA 20) and the triangular part of the inferior frontal gyrus (Tri-IFG) in the right hemispheres were also found.

Similarly, for the semantic processing of the Japanese sentence, the left-hemisphere regions were predominantly activated and partially overlapping with the semantic processing of the Chinese sentence (**Figure 3** and **Table 1B**). Clusters in the left hemisphere extended from the IOG (BA 18) to the fusiform gyrus (BA 37), from the precentral gyrus (BA 6) and the SMA (BA 6) to the Oper-IFG (BA 44), from the SPG (BA 7) to the MOG (BA 19), and the MTG (BA 21/22). In the right hemisphere, the clusters encompassed were the IOG (BA 19) to the lingual gyrus (BA 18) and the calcarine cortex (BA 17), as well as the AG (BA 7) to the MOG (BA 19), and the Tri-IFG.

In opposition to the Chinese and Japanese sentence processing, the semantic processing of the picture activated regions more bilaterally (**Figure 3** and **Table 1C**). Except for the cluster spreading from the superior temporal pole (BA 38) via the superior temporal gyrus (STG; BA 22) to the MTG (BA 21) in the right hemisphere, the regions that were activated both in the left and right hemispheres were symmetrical to some degree. The first two regions were the cluster extended from the internal SMA (BA 32) and the precentral gyrus (BA 6) to the Tri-IFG (BA 44) in the left hemisphere, and the cluster extended from the precentral gyrus (BA 6) and the middle frontal gyrus (MFG; BA 6) to the Tri-IFG in the right hemisphere. The other two regions were the broad areas located in the occipital lobe spreading from the left fusiform gyrus to the right hippocampus as the peak locations.

Furthermore, we observed the greater activity in the left inferior parietal gyrus (BA 40), the right supramarginal gyrus, and middle occipital gyrus for coherent compared with incoherent semantic processing (details are provided in the **Supplementary Material**).

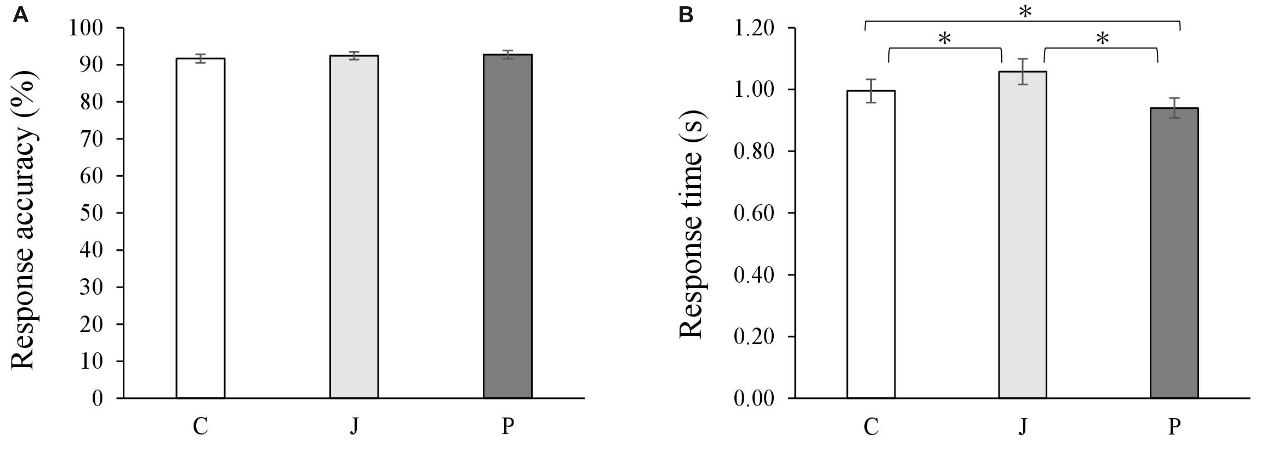


FIGURE 2 | Average accuracies **(A)** and response time **(B)** on judging coherence of stimuli of all conditions. Bars represent means with standard errors. C represents the Chinese sentence condition, J represents the Japanese sentence condition, and P represents the picture condition (*$p < 0.05$).

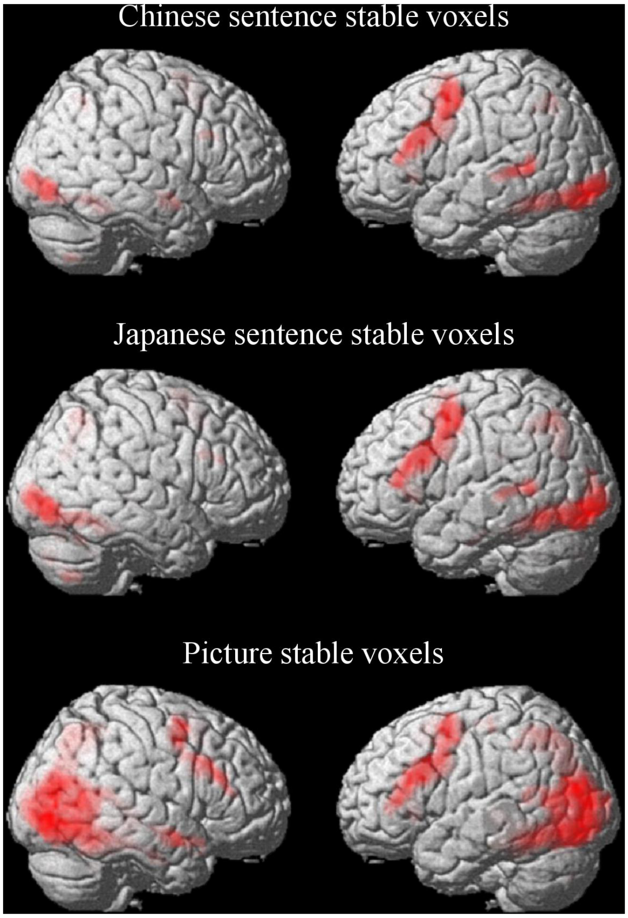


FIGURE 3 | Stable voxels for the Chinese sentence, Japanese sentence and picture in semantic processing. Stable voxel clusters significant at $p < 0.05$, FWE-corrected, extent threshold = 15 voxels.

Searchlight MVPA Analysis

Within-Language/Modality Classification

Figure 4A and **Table 2A** exhibit the areas in which the significant classification accuracies were found for the Chinese sentence. These regions were located in the left AG (BA 39) and extended to the MOG (BA 19). No significant classification accuracies were found within the Japanese sentence.

Areas involved in the classification of the pictures were more bilateral (**Figure 4B** and **Table 2B**) and included the left parieto-occipital regions spreading from the AG (BA 39) to the cuneus (BA 18/19), precuneus (BA 5), the lingual gyrus (BA 17/19), and the left MFG (BA46). The right MOG extending to the superior occipital gyrus (SOG; BA 7) and the precuneus (BA 7) were also noted.

Across-Language/Modality Classification

Significant across-language (i.e., Chinese sentence vs. Japanese sentence) classification accuracies were found in the left inferior parietal gyrus (IPG), which extends from the supramarginal gyrus (SMG; BA 40) to the AG (BA 39/7), and in the left precentral gyrus extending to the Oper and Tri-IFG (BA 44/45; **Figure 4C** and **Table 2C**).

Significant across-modality (i.e., Japanese sentence vs. picture) classification accuracy involved the left IPG extending from the SMG (BA 40) to the AG (BA 39/7) (**Figure 4D** and **Table 2D**). No significant classification accuracy was found between the Chinese sentence and the picture.

Results of using the TFCE also showed significant above chance classification accuracies for within- and across-language/modality classification. In the within-picture classification, significant classification accuracies were found in bilateral parieto-occipital regions (**Supplementary Figure 4A**). In the across-language classification, significant classification accuracies were found in the left IPG and the left IFG (**Supplementary Figure 4B**), though the regions significantly activated were smaller than those obtained using the searchlight MVPA. In contrast, significant classification accuracies were not observed for the within-Chinese sentences and the Japanese vs. picture classifications.

ROI Analysis

The mean classification accuracies in each ROI were contrasted with the chance level of accuracy (50%) using a one-sample *t*-test to accomplish the ROI analysis. For the within-Chinese sentence, the left BA 22 and BA 45 showed significant classification accuracies. For the within-Japanese sentence, the left BA 40, right BA 22, and BA 39 revealed significant classification accuracies. For the within-picture, the significant classification accuracies were shown in the bilateral BA 22, BA 39, and BA 40, the right BA 39, and the left BA 17, which was involved in primary visual information processing.

For the across-language classification, significant classification accuracies were shown in the bilateral BA 44, the left BA 39, BA 40, and BA 45. Across modalities (i.e., Japanese sentence vs. picture), significant accuracies were shown in the left BA 39 and BA 40, and in the right BA 45. In contrast to the searchlight analysis, the ROI analysis revealed significant cross-modality classification accuracies for the Chinese sentence vs. the picture in the bilateral BA 39 and BA 40 (**Figure 5**).

DISCUSSION

The present study used MVPA to investigate the common neural system of the semantic processing during sentence comprehension across languages in bilinguals. The significant classification accuracies indicate the existence of a common neural semantic representation in the higher language processing level. More specifically, the common neural representation was found to be situated in the left inferior parietal gyrus extending from the angular gyrus to the supramarginal gyrus, and the opercular and triangular part of the left inferior frontal gyrus. The results of this study also suggest that the left inferior parietal gyrus, in particular, the left angular gyrus and supramarginal gyrus, is pivotal to the processing of semantics regardless of the modality.

TABLE 1 | Activation for Chinese sentence **(A)**, Japanese sentence **(B)** and picture **(C)** semantic processing.

	Voxels	BA	T	MNI		
				x	y	z
(A) Chinese sentence						
L precentral gyrus	263	6	14.74	−42	2	47
L opercularis inferior frontal gyrus			12.04	−45	17	23
L opercularis inferior frontal gyrus		44	11.96	−45	5	29
L inferior occipital gyrus	565	18	14.37	−15	−94	−7
L fusiform gyrus		37	14.01	−45	−55	−16
L inferior occipital gyrus		18	13.21	−24	−94	−7
L supplementary motor area	204	6	12.36	−3	2	62
L supplementary motor area			11.81	−3	8	53
R middle cingulum gyrus			6.19	9	20	44
L middle temporal gyrus	132	21	11.37	−54	−49	8
L middle temporal gyrus		22	10.35	−57	−37	5
L insula	32	47	9.98	−30	23	2
L superior parietal gyrus	60	7	8.64	−27	−64	50
L cerebellum	21		7.79	−3	−52	−37
R fusiform gyrus	347	37	12.05	39	−49	−19
R lingual gyrus		18	11.94	18	−85	−7
R calcarine cortex		17	11.20	15	−91	−1
R cerebellum	18		10.58	30	−67	−52
R middle temporal gyrus	40	20	7.70	51	−10	−13
R triangularis inferior frontal gyrus	26		7.34	39	23	26
(B) Japanese sentence						
L inferior occipital gyrus	832	18	16.93	−15	−94	−7
L inferior occipital gyrus		18	14.48	−24	−94	−7
L fusiform gyrus		37	14.04	−42	−58	−13
L precentral gyrus	680	6	15.42	−42	2	47
L opercularis inferior frontal gyrus			12.64	−51	17	23
L precentral gyrus		44	12.39	−45	5	32
L supplementary motor area	22	6	12.16	−3	2	62
L supplementary motor area		32	11.10	−3	11	50
L insula	47	47	11.22	−30	23	2
L superior parietal gyrus	200	7	10.68	−27	−64	50
L middle occipital gyrus		19	7.87	−27	−70	29
L middle temporal gyrus	103	21	10.22	−54	−49	8
L middle temporal gyrus		22	9.09	−54	−37	5
L thalamus	15		8.34	−9	−13	5
L putamen	31		8.16	−21	2	5
R inferior occipital gyrus	675	19	13.70	30	−82	−13
R lingual gyrus		18	13.24	18	−82	−10
R calcarine cortex		17	12.52	15	−91	−1
R angular gyrus	99	7	9.90	30	−61	50
R middle occipital gyrus		19	8.50	33	−70	26
R cerebellum	27	37	9.12	30	−61	−28
R triangularis inferior frontal gyrus	30		7.02	42	26	26
(C) Picture						
L supplementary motor area	197	32	12.93	−3	14	50
L triangularis inferior frontal gyrus	675		12.35	−45	26	23
L opercularis inferior frontal gyrus		44	11.38	−45	5	29
L precentral gyrus		6	10.53	−42	5	44
L insula	36	47	9.93	−30	23	2
L postcentral gyrus	19		7.70	−39	−25	59
L cerebellum	16		7.54	−9	−73	−25

(Continued)

TABLE 1 | Continued

	Voxels	BA	T	MNI		
				x	y	z
L cerebellum	16		7.49	−9	−76	−40
R hippocampus	5103		20.97	24	−28	−4
L fusiform gyrus		37	18.56	−39	−58	−16
L fusiform gyrus		37	18.21	−27	−49	−10
R precentral gyrus	342	6	11.68	39	2	47
R triangularis inferior frontal gyrus			10.29	48	29	23
R middle frontal gyrus		6	9.73	36	2	56
R middle temporal gyrus	127	21	9.78	60	2	−16
R superior temporal gyrus		22	9.60	54	−7	−13
R superior temporal pole		38	9.10	48	14	−19
R amygdala	38	34	9.16	30	−1	−19
R hippocampus		35	9.00	21	−7	−19

Clusters of voxels significant at $p < 0.05$, FWE-corrected, extend threshold = 15 voxels. Region labels apply to the entire extent of the cluster with peak maxima designated by first locale cited.

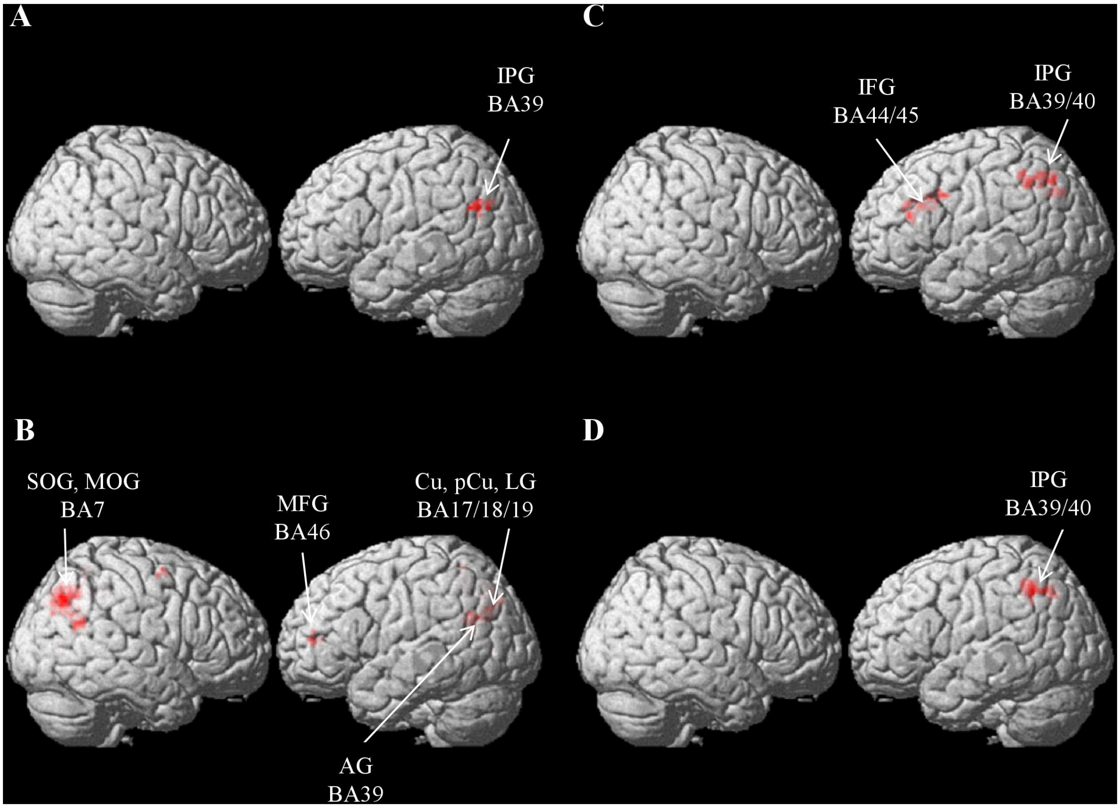


FIGURE 4 | Results of the searchlight analysis, showing significant classification accuracies above chance level (50%) from averaged whole-brain maps from all the participants at a cluster level FWE corrected at $p < 0.05$. **(A)** Results for within-Chinese classification. **(B)** Results for within-picture classification. **(C)** Results for across-language classification. **(D)** Results for across-modality classification.

Across Languages

The univariate analysis showed similar brain activation associated with the processing of the same sentence semantics for both Chinese and Japanese. This suggests that a common neural representation may exist across languages but could not allow the identification of the exact regions for which the MVPA was conducted.

Despite the inconsistent results revealed by previous studies investigating the common neural representation of word semantic processing across languages using the MVPA approach

TABLE 2 | Brain areas showed significant across-language **(A)**, across-modality **(B)**, within-Chinese **(C)** and within-picture **(D)** classification accuracies.

	Voxels	BA	T	MNI		
				x	y	z
(A) Across languages						
L inferior parietal gyrus	47	40	5.93	−42	−46	41
L inferior parietal gyrus		40	4.93	−48	−46	50
L angular gyrus		7	4.81	−36	−67	44
L precentral gyrus	36	44	5.07	−48	11	38
L triangularis inferior frontal gyrus		45	4.83	−45	32	20
L triangularis inferior frontal gyrus			4.61	−42	26	29
(B) Across Japanese and picture						
L inferior parietal gyrus	48	40	6.30	−45	−55	47
L angular gyrus		39	5.24	−39	−61	47
L inferior parietal gyrus		40	4.78	−42	−46	50
(C) Within Chinese						
L angular gyrus	22	39	4.43	−45	−61	29
L angular gyrus		39	4.41	−42	−58	26
L middle occipital gyrus		39	4.36	−42	−70	26
L angular gyrus		39	4.16	−48	−64	26
L angular gyrus		39	3.85	−48	−58	26
L middle temporal gyrus		39	3.82	−45	−64	20
L middle occipital gyrus		19	3.77	−36	−70	32
(D) Within picture						
L cuneus	16	19	6.41	−12	−79	38
L cuneus		18	3.65	−15	−73	32
L precuneus	14	5	5.19	−3	−52	62
L precuneus			4.74	−3	−58	47
L angular gyrus	13	39	4.60	−42	−55	23
L angular gyrus		39	4.50	−48	−61	29
			3.86	−39	−61	17
L lingual gyrus	11	17	4.60	−3	−64	8
L lingual gyrus		17	4.20	−9	−76	2
L calcarine cortex		17	3.69	−9	−70	8
L middle occipital gyrus	14	19	4.38	−33	−67	29
L middle frontal gyrus	11	46	4.12	−36	47	14
R middle occipital gyrus	200	19	7.84	39	−70	35
R superior occipital gyrus		7	7.64	30	−70	41
R precuneus		7	7.01	9	−73	41

Clusters of voxels significant at $p < 0.05$, FWE-corrected. Region labels apply to the entire extent of the cluster with peak maxima designated by first locale cited.

(Buchweitz et al., 2012; Correia et al., 2014; Van de Putte et al., 2017), it is possible to assume that the left temporoparietal conjunction regions are highly involved in the semantic processing of languages. Indeed, these regions were active when subjects were exposed to sentences or pictures depicting human events (Jouen et al., 2015). The present study observed significant classification accuracies in the left IPG, i.e., the AG (BA 39) and the SMG (BA 40), across languages for the semantic processing of sentences. Accordingly, the role of the left AG and SMG in the semantic processing of languages is suggested regardless of the processing level (word or sentence).

The semantics of languages are conveyed by symbols such as characters and/or sounds, which can be combined into words and/or sentences. Though the processes are divergent in the initial stages, the visual and auditory information must be mapped onto meanings to understand a sentence (Humphries et al., 2007). Thus, the neural pathways underpinning the visual and auditory semantic processing converge in the left AG (Bemis and Pylkkänen, 2013) for comprehension. Further, a series of further semantic processes are executed in the left AG to understand exactly the meanings conveyed by the words, especially by the sentences (Humphries et al., 2007). The most crucial process is adjusting the input information by verifying the already-existing knowledge. This manipulation may be executed in the left AG through retrieving the knowledge of the objects and events where it stored (Binder et al., 2009; Binder and Desai, 2011; Noonan et al., 2013). Then, the word semantic judging (Bonner et al., 2013) and naming and reading of the word (Seghier et al., 2010) could be executed. Constructing the meanings of the individual words (Price et al., 2016), the sentence could be comprehended (Pallier et al., 2011). Our findings reconfirm the established importance of the left AG in semantic processing, and implicate that the left AG is a critical region for semantic processing of languages transcending different languages and the processing levels. Meanwhile, the semantics conveyed by the sentences presented in the study concerned events in which entities interacted in space and time. Therefore, it also suggests that the left AG is undoubtedly necessary in the processing of event-related semantics (Binder and Desai, 2011; Seghier, 2013; Jouen et al., 2015; Baldassano et al., 2017).

The SMG, another part of the left IPG that is anterior to the AG, was also observed in significant classification accuracy across languages. This region is traditionally considered to underlie phonological processing such as phonological recognition, phonological control, and production (Booth et al., 2004; Prabhakaran et al., 2006). The SMG probably contributes preferentially to phonological aspects during word recognition. On the other hand, it is also reported to be critical for semantic processing (Stoeckel et al., 2009), especially in reading, which requires the recognition of visual stimuli and their linking to meanings (Sliwinska et al., 2012) as accomplished in our study. Further, the posterior part of the SMG is suggested to be the area where lexical and sublexical cues are integrated (Oberhuber et al., 2016), the lexical phonological retrieval is controlled and from the orthography to phonology is mapped (Price, 2018), and the meanings processed in the AG are bound to recognize the word (Lee et al., 2007). In the semantic processing of the sentence where words were formed, it is possible to assume that the continuous linking of lexical cues and meanings is required. Overall, as our study has indicated, the cooperation of both the left AG and SMG in the semantic processing of the sentence is demanded.

Another significant classification accuracy across languages was observed in divergence with previous MVPA studies of the semantic processing of words in the left inferior frontal gyrus, i.e., Oper-IFG (BA 44) and Tri-IFG (BA 45) which constitute Broca's area, a classical language processing area. Broca's area might underlie not only the language production but also various other language functions (Geschwind, 1970; Kim et al., 1997),

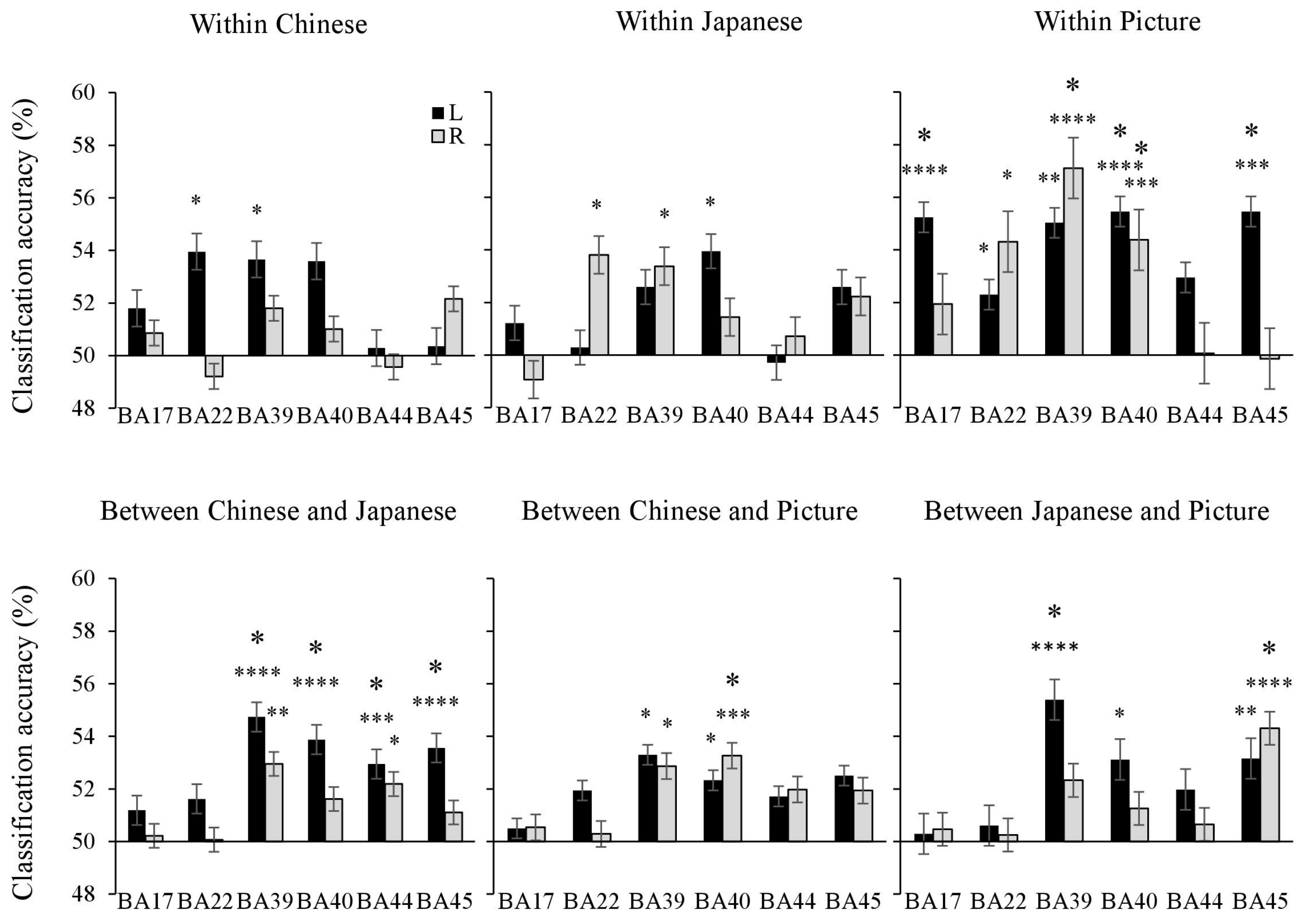


FIGURE 5 | The mean classification accuracies in each ROI were contrasted with the chance level of accuracy (50%) using a one-sample *t*-test. Bars represent standard errors. The small and large asterisks indicate statistical significance of *p* values uncorrected and corrected for multiple comparisons using Holm-Bonferroni procedure. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

one of them being the executive control of language. Apart from the comparatively simple processing such as the semantic processing of a sentence, the participants of this study were required to maintain the meanings of the first stimulus of the stimuli pair while processing the second stimulus to complete the evaluation task. Further, processes such as lexical retrieval and matching with previously held knowledge were needed to be executed simultaneously to accomplish the exact comprehension of the meanings. All the execution of these above processes is considered to be mandated by Broca's area (Devlin et al., 2003; Whitney et al., 2011; Noonan et al., 2013; Ralph et al., 2017).

Another role of Broca's area is the syntactic processing of language (Caplan, 2006; Grodzinsky and Santi, 2008; Friederici, 2011); BA 44 is especially considered to be the core region of the syntactic processing (Friederici, 2011, 2012) that provides strong cues for determining meanings (Humphries et al., 2007). As hypothesized, this syntactic processing-associated region was observed in the study. This may explain the discrepancy from the previous MVPA studies and indicate that the syntactic processing is critical for the sentence comprehension. Toward the determination of meanings of sentences, the syntactic information decoded by analyzing the constructions of the words and phrases which form the sentences will be mapped with the semantic information (Bookheimer, 2002; Friederici, 2011). Because of the preferential contribution of BA 44 to the syntactic processing, BA 44 is necessary to interact with the region that subserves the semantic processing, which is considered to be BA 45 (Friederici, 2011) to determine meanings. Based on the results of the present study, it is possible to assume that syntactic and semantic integration occur in Broca's area as Hagoort (2005, 2014) suggested. In this processing, the syntactic working memory is also required, which is considered to be one role of Broca's area (Fiebach et al., 2005; Makuuchi and Friederici, 2013). In the higher level of the semantic processing of language performed in our study, besides the semantic processing, sentence comprehension also demands both the executive control and the syntactic processing. Therefore, the involvement of Broca's area is presumable.

Broca's area presented more superior extension to the ventral part of the precentral gyrus (BA 6) in the present study.

As Hagoort (2005, 2014) suggested, it is more appropriate to refer to the BA 44, BA 45, BA 47, and BA 6 of the left inferior frontal cortex as Broca's area, because adjacent areas such as BA 47 and BA 6 are also involved in language processing. This finding of the present study is greatly consistent with Hagoort's assumption. Likewise, this finding might indicate the relevance of the motor-related system (i.e., BA 6) during comprehension of action-related sentences (e.g., Hauk et al., 2004; Hauk and Pulvermüller, 2011; Jouen et al., 2015), and the activation of acoustic representations during speech comprehension (Hickok and Poeppel, 2004, 2007; Cheung et al., 2016). Meanwhile, Broca's area is connected via the superior longitudinal fasciculus with the left AG and SMG. Hence, it makes sense that the robust neural representation associated with the higher level of semantic processing is situated in the left inferior parietal gyrus (i.e., the AG and the SMG) and the left inferior frontal gyrus (BA 44 and 45) (Horwitz et al., 1998; Frey et al., 2008; Kelly et al., 2010).

Across Modalities

A significant classification accuracy was observed in the left IPG (i.e., the AG and SMG) for semantic processing across modalities (i.e., Japanese sentence vs. picture and Chinese sentence vs. Japanese sentence). This result revealed the modality-independent common neural representation. The univariate analysis of the coherence judgment (see **Supplementary Material**) also showed the involvement of the left SMG for coherent semantic processing regardless of the modalities. These findings tend to support the idea proposed by Damasio (1989) and Meyer and Damasio (2009) that there are convergence zones where the features associated with different objects and events and/or information conveyed by different sensory systems are bound. The features and/or information were considered to be the processing of the meanings of the features and/or the information (Mahon and Caramazza, 2008). Despite the fact that the neural basis of the convergence zone is still controversial, the association of the temporoparietal regions overlapped to some degree (Jefferies, 2013; Simanova et al., 2014; Jouen et al., 2015; Wurm and Caramazza, 2019). Specifically, it is posited that the inferior parietal lobe, the ventral and lateral temporal lobes are involved in the higher-level convergence processing where the binding representation from multiple modalities encode an abstract or schematic concept (Binder and Desai, 2011; Simanova et al., 2014). Sentence comprehension requiring fluent conceptual combinations as in the present study demands the higher-level convergence processing of complex information. The information from the languages and modalities needs to be integrated with the stored knowledge in the convergence zone, which is the left IPG (Lau et al., 2008; Binder et al., 2009; Binder and Desai, 2011; Bonner et al., 2013; Seghier, 2013). Though the searchlight analysis evidenced the absence of significant classification accuracy between the Chinese sentence and the picture, the ROI analysis showed significant classification accuracy in the left BA 39 and BA 40. Hence, the results of the present investigation make it possible to indicate that the left inferior parietal gyrus (BA 39 and 40) is a modality-independent convergence zone for higher semantic processing.

CONCLUSION

This study aimed to investigate whether an across-language sentence comprehension system exists using MVPA with Chinese–Japanese bilinguals, and whether such a system shares a common foundation for the broader comprehension of meaning in images. The results first suggest that the existence of the common neural system across languages in the semantic processing of sentences is located in the left inferior parietal gyrus (BA 39 and BA 40) and in the left inferior frontal gyrus (BA 44 and BA 45), which is also known as Broca's area. Second, the findings elucidate the specific functioning of the left inferior parietal gyrus as a modality-independent convergence zone, particularly in higher semantic processing as required for understanding sentences and images.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Hokkaido University Institutional Review Board with written informed consent from all subjects. All subjects provided written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Hokkaido University Institutional Review Board.

AUTHOR CONTRIBUTIONS

ZH and KO conceived and designed the experiment. ZH performed the experiment, analyzed the data, and drafted the manuscript. HY and SN coordinated the data analysis. YY assisted in performing the experiment and collected the data. CM-L, JV-D, and PD contributed to conceptualization of the experiment and provided the material. PD reviewed the manuscript. KO supervised the experiment and the data analysis, reviewed and revised the manuscript.

FUNDING

This work was supported by Graduate Grant Program of the Graduate School of Letters, Hokkaido University, Japan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00380/full#supplementary-material

REFERENCES

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron* 95, 709–721. doi: 10.1016/j.neuron.2017.06.041

Bemis, D. K., and Pylkkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb. Cortex* 23, 1859–1873. doi: 10.1093/cercor/bhs170

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Bonner, M. F., Peelle, J. E., Cook, P. A., and Grossman, M. (2013). Heteromodal conceptual processing in the angular gyrus. *Neuroimage* 71, 175–186. doi: 10.1016/j.neuroimage.2013.01.006

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188. doi: 10.1146/annurev.neuro.25.112701.142946

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2004). Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* 16, 1234–1249. doi: 10.1162/0898929041920496

Buchweitz, A., Mason, R. A., Hasegawa, M., and Just, M. A. (2009). Japanese and English sentence reading comprehension and writing systems: an fMRI study of first and second language effects on brain activation. *Biling Lang. Cogn.* 12, 141–151. doi: 10.1017/S1366728908003970

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., and Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain Lang.* 120, 282–289. doi: 10.1016/j.bandl.2011.09.003

Caplan, D. (2006). Why is broca's area involved in syntax. *Cortex* 42, 469–471. doi: 10.1016/S0010-9452(08)70379-4

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16. doi: 10.1002/ana.20319

Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 27:1–27:27. doi: 10.1145/1961189.1961199

Cheung, C., Hamiton, L. S., Johnson, K., and Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife* 5:e12577. doi: 10.7554/eLife.12577.001

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *J. Neurosci.* 34, 332–338. doi: 10.1523/JNEUROSCI.1302-13.2014

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62. doi: 10.1016/0010-0277(89)90005-X

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. S. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15, 71–84. doi: 10.1162/089892903321107837

Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., and Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Hum. Brain Mapp.* 24, 79–91. doi: 10.1002/hbm.20070

Frey, S., Campbell, J. S. W., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444. doi: 10.1523/JNEUROSCI.2388-08.2008

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Psychol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D., Frackowiak, R. S. J., et al. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Fromkin, V., Rodman, R., and Hyams, N. (2014). *An Introduction to Language*, 10th Edn. Boston: Wadsworth.

Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.

Grodzinsky, Y., and Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Curr. Opin. Neurobiol.* 16, 240–246. doi: 10.1016/j.conb.2006.03.007

Grodzinsky, Y., and Santi, A. (2008). The battle for Broca's region. *Trends Cogn. Sci.* 12, 474–480. doi: 10.1016/j.tics.2008.09.001

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr. Opin. Neurobiol.* 28, 136–141. doi: 10.1016/j.conb.2014.07.013

Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neurosci. Biobehav. Rev.* 81, 194–204. doi: 10.1016/j.neubiorev.2017.01.048

Hauk, O., Johnsrude, I., and Pulvermü, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9

Hauk, O., and Pulvermüller, F. (2011). The lateralization of motor cortex activation to action-words. *Front. Hum. Neurosci.* 5:149. doi: 10.3389/fnhum.2011.00149

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Horwitz, B., Rumsey, J. M., and Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8939–8944. doi: 10.1073/pnas.95.15.8939

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage* 36, 924–932. doi: 10.1016/j.neuroimage.2007.03.059

Ingram, J. C. L. (2007). *Neurolinguistics*, 1st Edn, New York, NY: Cambridge University Press.

Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* 49, 611–625. doi: 10.1016/j.cortex.2012.10.008

Jouen, A. L., Ellmore, T. M., Madden, C. J., Pallier, C., Dominey, P. F., and Ventre-Dominey, J. (2015). Beyond the word and image: characteristics of a common meaning system for language and vision revealed by functional and structural imaging. *Neuroimage* 106, 72–85. doi: 10.1016/j.neuroimage.2014.11.024

Keller, T. A., Carpenter, P. A., and Just, M. A. (2001). The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cereb. Cortex* 11, 223–237. doi: 10.1093/cercor/11.3.223

Kelly, C., Uddin, L. Q., Shehzad, Z., Margulies, D. S., Castellanos, F. X., Milham, M. P., et al. (2010). Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. *Eur. J. Neurosci.* 32, 383–398. doi: 10.1111/j.1460-9568.2010.07279.x

Kim, K. H. S., Relkin, N. R., Lee, K.-M., and Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature* 388, 171–174. doi: 10.1038/40623

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* 13, 3863–3868. doi: 10.1073/pnas.0600244103

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lee, H., Devlin, J. T., Shakeshaft, C., Stewart, L. H., Brennan, A., Glensman, J., et al. (2007). Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* 27, 1184–1189. doi: 10.1523/JNEUROSCI.4442-06.2007

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004

Makuuchi, M., and Friederici, A. D. (2013). Hierarchical functional connectivity between the core language system and the working memory system. *Cortex* 49, 2416–2423. doi: 10.1016/j.cortex.2013.01.007

Marslen-Wilson, W., and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition* 8, 1–71. doi: 10.1016/0010-0277(80)90015-3

Meyer, K., and Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32, 376–382. doi: 10.1016/j.tins.2009.04.002

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191–1195. doi: 10.1126/science.1152876

National Institute for Japanese Language, and Linguistics (2018). *The Balanced Corpus of Contemporary Written Japanese*. Tokyo: National Institute for Japanese Language, and Linguistics.

Newman, S. D., Ikuta, T., and Burns, T. (2010). The effect of semantic relatedness on syntactic analysis: an fMRI study. *Brain Lang.* 113, 51–58. doi: 10.1016/j.bandl.2010.02.001

Nichols, T. E., and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25. doi: 10.1016/B978-012264841-0/50048-2

Noonan, K. A., Jefferies, E., Visser, M., and Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J. Cogn. Neurosci.* 25, 1824–1850. doi: 10.1162/jocn_a_00442

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., et al. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cereb. Cortex* 26, 4212–4226. doi: 10.1093/cercor/bhw251

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527. doi: 10.1073/pnas.1018711108

Prabhakaran, R., Blumstein, S. E., Myers, E. B., Hutchison, E., and Britton, B. (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia* 44, 2209–2221. doi: 10.1016/j.neuropsychologia.2006.05.025

Price, A. R., Peelle, J. E., Bonner, M. F., Grossman, M., and Hamilton, R. H. (2016). Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation. *J. Neurosci.* 36, 3829–3838. doi: 10.1523/jneurosci.3120-15.2016

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88. doi: 10.1111/j.1749-6632.2010.05444

Price, C. J. (2018). The evolution of cognitive models: from neuropsychology to neuroimaging and back. *Cortex* 107, 37–49. doi: 10.1016/j.cortex.2017.12.020

Ralph, M. A. L., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *Neurosci.* 19, 43–61. doi: 10.1177/1073858412440596

Seghier, M. L., Fagan, E., and Price, C. J. (2010). Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J. Neurosci.* 30, 16809–16817. doi: 10.1523/JNEUROSCI.3377-10.2010

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., and Just, M. A. (2011). Commonality of neural representations of words and pictures. *Neuroimage* 54, 2418–2425. doi: 10.1016/j.neuroimage.2010.10.042

Simanova, I., Hagoort, P., Oostenveld, R., and Van Gerven, M. A. J. (2014). Modality-independent decoding of semantic information from the human brain. *Cereb. Cortex* 24, 426–434. doi: 10.1093/cercor/bhs324

Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., and Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Front. Psychol.* 3:161. doi: 10.3389/fpsyg.2012.00161

Smith, S. M., and Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98. doi: 10.1016/j.neuroimage.2008.03.061

Stoeckel, C., Gough, P. M., Watkins, K. E., and Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex* 45, 1091–1096. doi: 10.1016/j.cortex.2008.12.004

Tan, L. H., Laird, A. R., Li, K., and Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: a meta-analysis. *Hum. Brain Mapp.* 25, 83–91. doi: 10.1002/hbm.20134

Van de Putte, E., De Baene, W., Brass, M., and Duyck, W. (2017). Neural overlap of L1 and L2 semantic representations in speech: a decoding approach. *Neuroimage* 162, 106–116. doi: 10.1016/j.neuroimage.2017.08.082

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., and Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb. Cortex* 21, 1066–1075. doi: 10.1093/cercor/bhq180

Wurm, M. F., and Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nat. Commun.* 10:289. doi: 10.1038/s41467-018-08084-y

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., and Just, M. A. (2017a). Commonalities and differences in the neural representations of English, Portuguese, and Mandarin sentences: when knowledge of the brain-language mappings for two languages is better than one. *Brain Lang.* 175, 77–85. doi: 10.1016/j.bandl.2017.09.007

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., and Just, M. A. (2017b). Commonality of neural representations of sentences across languages: predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *Neuroimage* 146, 658–666. doi: 10.1016/j.neuroimage.2016.10.029

Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., and Raizada, R. D. S. (2016). Semantic structural alignment of neural representational spaces enables translation between english and Chinese Words. *J. Cogn. Neurosci.* 28, 1749–1759. doi: 10.1162/jocn_a_01000

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 14 November 2019
doi: 10.3389/fnhum.2019.00390

Children With Reading Difficulty Rely on Unimodal Neural Processing for Phonemic Awareness

Melissa Randazzo[1]*, Emma B. Greenspon[2,3], James R. Booth[4] and Chris McNorgan[2]

[1]Department of Communication Sciences and Disorders, Adelphi University, Garden City, NY, United States, [2]Department of Psychology, State University at Buffalo, New York, NY, United States, [3]Department of Psychology, Monmouth University, New Jersey, NJ, United States, [4]Department of Psychology and Human Development, Vanderbilt University, Tennessee, TN, United States

Phonological awareness skills in children with reading difficulty (RD) may reflect impaired automatic integration of orthographic and phonological representations. However, little is known about the underlying neural mechanisms involved in phonological awareness for children with RD. Eighteen children with RD, ages 9–13, participated in a functional magnetic resonance imaging (fMRI) study designed to assess the relationship of two constructs of phonological awareness, phoneme synthesis, and phoneme analysis, with crossmodal rhyme judgment. Participants completed a rhyme judgment task presented in two modality conditions; unimodal auditory only and crossmodal audiovisual. Measures of phonological awareness were correlated with unimodal, but not crossmodal, lexical processing. Moreover, these relationships were found only in unisensory brain regions, and not in multisensory brain areas. The results of this study suggest that children with RD rely on unimodal representations and unisensory brain areas, and provide insight into the role of phonemic awareness in mapping between auditory and visual modalities during literacy acquisition.

Keywords: reading difficulty, crossmodal integration, phonemic awareness, audiovisual integration, fMRI—functional magnetic resonance imaging, dyslexia

Edited by:
Xiaolin Zhou,
Peking University, China

Reviewed by:
Taomei Guo,
Beijing Normal University, China
Camila Rosa De Oliveira,
Faculdade Meridional (IMED), Brazil

Correspondence:
Melissa Randazzo
mrandazzo@adelphi.edu

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *11 April 2019*
Accepted: *21 October 2019*
Published: *14 November 2019*

Citation:
Randazzo M, Greenspon EB, Booth JR and McNorgan C (2019) Children With Reading Difficulty Rely on Unimodal Neural Processing for Phonemic Awareness. Front. Hum. Neurosci. 13:390. doi: 10.3389/fnhum.2019.00390

INTRODUCTION

Phonological awareness skills are important in learning to read in alphabetic languages (Bus and van IJzendoorn, 1999; Ehri et al., 2001; Lonigan et al., 2009). Phonemic awareness skills are metalinguistic skills that include the ability to manipulate the sound structure of oral language (Wagner and Torgesen, 1987; Share, 1995; Ehri et al., 2001; Ziegler and Goswami, 2005). These skills are a subset of phonological awareness skills that operate on smaller phonological segments at the level of the individual speech sound, or phoneme (Anthony et al., 2003; Ziegler and Goswami, 2005). Phonemic awareness is strongly associated with word reading and is more highly correlated with reading skills than rhyme-level awareness (Melby-Lervåg et al., 2012). Phonemic awareness, however, can be further subdivided into two complementary skills: synthesis and analysis (Perfetti et al., 1987). Synthesis refers to the ability to combine isolated phonemes into syllables or words. For example, blending phonemes /k/, /æ/, and /t/ into the word "cat" demonstrates the synthesis principle. Analysis refers to the ability to break words or syllables into smaller speech segments. For example, in elision tasks, given the instruction to produce "cat" without the "/k/," the verbal response "/æt/," demonstrates the analysis principle.

Component Constructs of Phonological Awareness

Synthesis and analysis are often considered a unified construct in studies examining the development of phonological awareness (Anthony and Lonigan, 2004; Lonigan et al., 2009). Several studies, however, suggest that these two skills are distinct aspects of phonemic awareness that are performed by divergent manipulations. Phoneme synthesis and analysis skills appear to have different developmental trajectories and play different roles in the reading acquisition process. Phoneme synthesis, which is the phonemic awareness skill underlying blending tasks, develops before phoneme analysis, which is the phonemic awareness skill underlying tasks requiring segmentation (Anthony et al., 2003; Lonigan et al., 2009). Different developmental trajectories suggest there is at least a partial dissociation between the contributions of these two skills during reading development. Studies examining the relationship between phoneme synthesis and analysis skills with early reading development and instructional outcomes have reinforced the notion that these constructs differentially contribute to reading performance. For example, in their study of preschoolers prior to formal reading instruction, Burgess and Lonigan (1998) found that letter-sound knowledge was a unique predictor of phoneme analysis in elision tasks but not of synthesis in blending tasks. Similarly, Kroese et al. (2000) found that elision along with phoneme reversal were strongest predictors of reading and spelling ability in children in late elementary school, with blending predicting less of the variance. These findings support both the arguments that analysis and synthesis are constructs that tap different aspects of phonemic awareness, and that a bidirectional relationship exists between phonemic awareness and reading skill.

An earlier attempt to more precisely characterize this relationship was carried out by Perfetti et al. (1987), who examined the unique contributions of synthesis and analysis skills towards reading development between first and second grade. This longitudinal study followed first graders exposed to either of two types of reading instruction emphasizing direct (i.e., phonics) or indirect (i.e., whole-word) decoding. The study demonstrated that the relationship between these skills and reading ability depended on the type of reading instruction administered, suggesting that these two phonemic awareness skills are differentially related to different reading strategies, and thus represent distinct skills. The authors concluded that synthesis skills enable reading development, presumably through bootstrapping orthographic assembly from existing skill in phonological assembly. However, acquired reading skill enables later analysis skills, thus providing an account in which phonemic awareness skills and reading development are mutually dependent.

A prominent model of reading acquisition has argued that phonemic-level awareness is a result of increased sensitivity to phonemes by exposure to orthography (Ziegler and Goswami, 2005), and consequently plays a crucial role in bridging phonological representation to orthographic input. This idea is consistent with the findings described above showing a bidirectional relationship between phoneme analysis skills and typically-developing reading ability, and is supported by numerous studies showing that orthographic knowledge influences phonological processing (Ehri and Wilce, 1980; Stuart, 1990; Castles et al., 2003; Desroches et al., 2010). Upon acquiring letter knowledge, readers may utilize this information on phoneme judgment tasks. Given even relatively brief exposure to orthographic representations, preschoolers also demonstrate a similar influence of orthography on phonemic awareness tasks (Castles et al., 2011). This suggests that a crossmodal influence of orthographic and phonological representations may accompany the learning of the alphabetic principle, and continue as a child learns to read. This is consistent with the argument that phoneme analysis skills are an experience-dependent outcome of skilled reading in opaque orthography (Mann and Wimmer, 2002).

Crossmodal Processing and Phonemic Awareness

Interventions for phonological awareness skills are typically motivated by the assumption that reading difficulty (RD) arises from deficits in auditory processing or phonological representation (Tallal, 1980; Shaywitz and Shaywitz, 2005). However, the role of orthography in the development of phonemic awareness skills and the reciprocal nature of phonemic awareness and reading ability suggest that a failure to integrate letters and speech sounds may contribute substantially towards RD. It has been suggested that decreased phonological awareness in children with RD may reflect impaired automatic integration of orthographic and phonological representations, signifying a crossmodal deficit in integrating visual letters and auditory speech sounds, rather than a unimodal auditory processing deficit (Vaessen et al., 2009). Deficits in letter-sound integration may be attributed to decreased automatization of pairing these associations (Bakos et al., 2017). Multisensory interactions provide useful constraints on lexical activations in either modality to the extent that they help disambiguate multiple competing representations (Seidenberg and McClelland, 1989; Harm and Seidenberg, 1999). Moreover, training such mapping between modalities, in turn, drives visual specialization (Fraga González et al., 2017) and can improve reading fluency (Žarić et al., 2015). We take the well-supported position that developmental reading difficulties arise in large part from an inability to accurately and quickly map between phonological and orthographic representations, or between auditory and visual modalities (Booth et al., 2004, 2007; Cao et al., 2006; Bitan et al., 2007).

Because they identify the neural correlates of cognitive processing, neuroimaging, and neurophysiological studies have contributed much towards our understanding of the interacting systems involved in typical and disordered reading. Neuroimaging and neurophysiological studies examining audiovisual integration of letters and speech sounds in children suggest that RD may be partly attributable to difficulties in crossmodal integration of these entities. In this view, crossmodal associations between the visual letter and auditory speech sound are reinforced through reading experience, which then refines phonemic awareness skills in typically developing (TD) readers (Blau et al., 2009). Blau et al. (2010) demonstrated enhanced letter-sound integration in audiovisual conditions for TD compared to dyslexic readers in a series of functional magnetic resonance imaging (fMRI) studies. These studies collectively identified an audiovisual integration network that includes brain regions actively engaged in both unimodal and crossmodal processing of phonological and orthographic representations.

Phonological processing and representation are critically supported by the primary auditory cortex (or Heschl's gyrus) and superior temporal gyrus (STG; Hickok and Poeppel, 2000; Humphries et al., 2014). Though coarsely characterized as a unimodal region, a posterior sub-region in the STG called the Planum Temporale (PT) and the posterior superior temporal sulcus (pSTS), both of which are anatomically proximal to visuomotor processing pathways, have been implicated as audiovisual integration sites in multiple domains (Calvert, 2001; Beauchamp et al., 2004; Stevenson and James, 2009; van Atteveldt et al., 2009). Consistent with a model in which RD arises in part from disordered audiovisual integration, crossmodal activation within PT and pSTS has been shown to differentiate TD readers and children with RD (Blau et al., 2010). In this neuroanatomical model of audiovisual letter to sound mapping, orthographic representations mediate the auditory cortical response to speech sounds in the PT only for TD readers. RD readers, in contrast, under-activate unisensory processing regions to speech sounds and visual letters in the STG and fusiform gyrus (FG), respectively, possibly resulting from deficient crossmodal mediation. Consistent with this model, Blau and colleagues found that the visual response to print in the FG was associated with crossmodal processing effects in the PT, mediated by reading skill. Price and Devlin's (2011) Interactive Account similarly argues that audiovisual integration plays a critical mediating role in reading development, claiming that the specialization of FG for orthographic processing is a consequence of internally-driven (i.e., top-down) phonological input facilitating the perceptually-driven (i.e., bottom-up) visual object processing system. This facilitation hinges on effective crossmodal integration. The introduction of additional disambiguating information helps reduce uncertainty and identifies the most probable lexical representation. It follows that ineffective crossmodal integration may provide no useful information, or even misleading information. Uncovering and improving how children with RD cope with this challenge is thus a central goal for those who research and work with these populations. Collectively, this body of literature suggests that typical reading development relies on successful audiovisual integration and that RD is associated with reduced integration between modalities (Richlan, 2019).

How and When do Children With Reading Difficulty Make Use of Phonemic Awareness?

Given the body of research implicating disrupted audiovisual integration in RD, and that remediation for poor readers is often focused on assessment and improvement of phonemic awareness, it is important to understand the neural mechanisms underlying distinct aspects of phonemic awareness, and how they interact with those underlying audiovisual integration. Little is known about the underlying neural mechanisms involved in phonological awareness for children with RD in deep orthographies; indeed, rectifying this gap in the literature partly motivates the present research.

Frost et al. (2009) examined the relationship between phonemic awareness and brain activation for print and speech in TD readers and found that a composite measure of higher phonemic awareness skill was associated with increased activation in the left STG for the processing of print. Although the study examined print and speech processing in separate modality conditions, phoneme analysis was more sensitive to the overlap of print and speech processing than phoneme synthesis. This suggests that phoneme analysis skills are more related to the audiovisual integration of speech and print processing than are phoneme synthesis skills.

The position that normal and disordered reading differentially engages audiovisual integration processes is supported by an fMRI study, which found that the connection between audiovisual integration and reading skill differed for TD and RD readers (McNorgan et al., 2013). This study showed that Elision skill was related to neural activity when engaged in audiovisual processing, but not auditory- or visual-only processing. Moreover, this relationship held in TD but not children with RD, even though TD and RD groups had overlapping Elision scores. This relationship between phoneme analysis and audiovisual lexical processing was driven by sensitivity to orthographic congruency in the FG and pSTS, regions strongly associated with orthographic processing (Tsapkini and Rapp, 2010) and audiovisual integration (Calvert, 2001), respectively. Similarly, Gullick and Booth (2014) found that pSTS activity is related to functional connectivity in the arcuate fasciculus, a tract that is related to individual differences in reading skill, during crossmodal rhyme judgment in typical readers. Broader consideration of the body of work on these regions comprising a crossmodal reading network suggests that phoneme analysis is related to audiovisual integration processes in TD children and that RD is associated with a breakdown in this relationship.

The purpose of the current study is to examine how phonemic awareness supports online rhyme judgment, a phonologically-based lexical task, in children with RD. The null effects associated with their RD sample necessarily lead McNorgan et al. (2013) to conclude there was no evidence of any relationship between neural processing dynamics and phonemic awareness skills in their RD sample. This consequently limited their framing of RD to the neurotypical processes in which they do not engage during reading, leaving a gap in our understanding of the neural correlates of phonological processing during reading in children with RD. It remains unclear which phonological processes children with RD do engage while reading, and how this processing relates to RD severity. We take advantage of our experimental design to examine how constructs of phonemic awareness are related to the degree of impairment and the magnitude of brain activity under different modality conditions. As phoneme analysis involves advancing abilities over time, we explored the influence of modality in rhyme judgment in children ages 9–13 years old who have received several years of reading instruction and span a continuum of reading ability.

We focused our analysis on a sub-network of left hemisphere regions for which the neuroimaging literature has shown consensus as being involved in phonological and orthographic processing and in audiovisual integration, and has been explicitly implicated in the models reviewed above. The inferior parietal lobule (IPL) is implicated in phonological processing and mapping between orthographic and phonological representations (Bitan et al., 2007; Moll et al., 2014). The FG is recognized for its specialization in print processing in skilled readers (Shaywitz et al., 2002; McCandliss et al., 2003; Dehaene and Cohen, 2011). The pSTS is widely regarded as an audiovisual integration site across domains, with a specific role in the integration of letters and speech sounds (Calvert, 2001; van Atteveldt et al., 2009; Blau et al., 2010). Finally, the STG contains associative auditory cortex and plays a critical role in processing phonological word forms (Pugh et al., 2001; Booth et al., 2002; Friederici, 2012).

Investigations of audiovisual integration depend on stimulus congruency, or the correspondence between representations, as this demonstrates how the processing of one representation influences the processing of the other. In studies investigating audiovisual integration in reading, congruency is examined between unimodal and crossmodal presentations at a small grain size (i.e., letters and speech-sounds; Froyen et al., 2008) or lexical rhyme judgment at a larger grain size (McNorgan et al., 2013, 2014). Given the inconsistency of the English orthography at the smaller grain sizes (e.g., letters), large grain sizes (e.g., words, syllables or rimes) play a greater role in early reading development because they provide greater consistency (Ziegler and Goswami, 2005) and are sensitive to skill-related differences in audiovisual lexical processing (Kronschnabel et al., 2014). Fluent reading in English necessitates the processing of larger grain sizes because the processing of smaller grain sizes utilizing a letter-by-letter decoding strategy will only be successful with words that have consistent grapheme to phoneme correspondences. Therefore, we assessed the neural response to inter-stimulus phonological congruency for unimodal (auditory-only) and crossmodal (audiovisual) items at the lexical level in a rhyme judgment task.

The body of research indicating that RD arises from a failure to integrate letters and speech sounds suggests that children with RD might favor unimodal processing of lexical items. Accordingly, we hypothesized phonemic awareness tasks in these children would draw on unimodal processing, rather than crossmodal integration, and thus that these effects would be evident in STG and FG, the two nodes in our reading network most strongly associated with unimodal processing of phonology and orthography, respectively. Moreover, given that previous research suggests differential relationships between phoneme analysis and synthesis skills and reading development, we were also interested in whether these skills bear different relationships to brain activation when analyzed within the context of unimodal (auditory-only) and crossmodal (audio-visual) presentation modalities in children along a continuum of reading ability. Because McNorgan et al. (2013) found Elision scores to be unrelated to neural processing under audio-visual presentation conditions in a sample of children with RD closely matched with a TD sample, we anticipated replicating this finding but hypothesized that phonemic awareness, which is associated with crossmodal processing in TD readers, would be predicted by unimodal processing in our larger sample.

MATERIALS AND METHODS

Participants

Participants were selected from among a larger group involved in a study investigating reading development in children with a range of reading ability. All participants were native English speakers, right-handed, had a normal or corrected-to-normal vision, and had no history of psychiatric illness, neurological disease or attention deficit hyperactivity disorder (ADHD). Study participants were recruited from the Chicago metropolitan area. Informed consent was obtained from participants and their parents. The Institutional Review Board at Northwestern University approved all procedures.

Participants with a prior diagnosis of RD were referred for the study. RD was quantified prior to admission to the study as a standard screening procedure. We evaluated non-verbal IQ using the *Wechsler Abbreviated Scale of Intelligence—Second Edition* (WASI-II; Wechsler, 2011) and reading skill using the Word Identification, Word Attack and Reading Fluency subtests of the *Woodcock-Johnson Tests of Achievement—III* (WJ III; Woodcock et al., 2001) and the Sight Word Efficiency (SWE) and Phonetic Decoding Efficiency (PDE) subtests of the *Test of Word Reading Efficiency* (TOWRE; Torgesen et al., 1999). Eighteen children met our eligibility requirements (11 males; mean age = 11 years, 8 months; range = 9 years, 10 months to 13 years, 11 months). Participants were included in the study if, in addition to presenting with a clinical diagnosis of RD, at least two of the five scaled scores were less than or equal to 95, at least one score at or below 91, and the mean of all scaled scores was less than or equal to 95. Participants had an average of 3.9 out of 5 standard scores below 90. All other scores fell in the average to below-average range across participants. These selection criteria enabled our correlational design to investigate reading skills in otherwise cognitively typical children diagnosed with RD but demonstrating a range of skills. Each participant's phonemic awareness was measured by the Elision and Blending subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999). Scores reflect the number of correct Elision or Blending transformations on a set of 20 progressively more difficult target items. Participants had near-average performance for the Blending and Elision measures of Phonological Awareness (group mean raw score Blending = 9.0, group mean raw score Elision 8.7, test mean raw score = 10). Participants had better than chance performance on the experimental task ($M = 0.65$, $SD = 0.11$) and without evidence of response bias across all scanning sessions,

TABLE 1 | Participant characteristics.

Measure	Mean Score (SD)	Range
WASI Performance IQ	100 (13)	74–127
WJ-III Word ID	90 (7)	67–113
WJ-III Word Attack	92 (5)	83–103
WJ-III Reading	90 (10)	67–113
TOWRE SWE	89 (10)	60–113
TOWRE PDE	88 (10)	71–104
CTOPP Elision	8.7 (2.8)	4–12
CTOPP Blending	9.0 (2.2)	5–14

Note: WASI, Wechsler Abbreviated Scale of Intelligence; WJ-III, Woodcock-Johnson Tests of Achievement—III; TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency; PDE, Phonetic Decoding Efficiency; CTOPP, Comprehensive Test of Phonological Processing. CTOPP raw scores reported above; CTOPP Elision score max = 20, test mean = 10; CTOPP Blending score max = 20, test mean = 10.

as measured by a d-prime analysis of responses (M = 0.64). Summary statistics for these participant characteristics appear in **Table 1**. We report raw, rather than scaled, CTOPP scores because they were used as dependent measures in the regression analyses that follow.

Experimental Procedure

Rhyme Judgment Task

On each trial, participants were presented with paired stimuli, the order of which was counterbalanced across participants in an event-related paradigm. For each scanning session, stimuli were presented in one of two modality conditions: In the cross-modal auditory/visual (AV) condition, the first item was presented auditorily and the second was presented visually. In the unimodal auditory/auditory (AA) condition, both items were presented in the auditory modality. Half the pairs of stimuli rhymed and half did not, and participants were asked to make a rhyme judgment response by pressing one of two keys on a handheld keypad. Participants were asked to respond as quickly and as accurately as possible, using their right index finger for a yes (rhyme) response and their right middle finger for a no (non-rhyme) response. Participants participated in two runs for each modality condition, each lasting approximately 7 min. Participants generally saw the AV condition followed by the AA condition, though this varied across participants as factors such as task accuracy and movement necessitated reacquiring data.

Figure 1 illustrates the design of each trial. Each stimulus item was presented for 800 ms, separated by a 200 ms interstimulus interval. Participants were free to respond as soon as the second stimulus item was presented. A red cross appeared for 2,200 ms following the presentation of the second word, signaling to the participant to respond if they had not already done so. Responses made after the red cross disappeared from the screen were not recorded and were counted as errors.

We examine rhyme judgment in orthographically consistent (O+) and orthographically inconsistent (O−) conditions, which we crossed with the rhyming (P+) and non-rhyming (P−) lexical conditions. Thus, the lexical trials included two rhyming conditions, comprising orthographically consistent (O+P+ e.g., CAGE RAGE) and orthographically inconsistent (O−P+ e.g., GRADE LAID) pairs. The lexical trials also included two non-rhyming conditions, comprising orthographically consistent (O+P−; e.g., SMART WART) and orthographically inconsistent (O−P−; e.g., TRIAL FALL) pairs. There were 12 trials of each rhyming condition in each run. All words were monosyllabic, having neither homophones nor homographs, and were matched across conditions for written word frequency in children (Zeno et al., 1995) and the sum of their written bigram frequency (English Lexicon Project[1]). The stimuli had a mean age of acquisition of 6.6 (SD = 1.9) years (Kuperman et al., 2012) and we thus expected participants to be generally familiar with the experimental stimuli. We restricted our analyses to the two rhyming conditions (i.e., those associated with "yes" responses) to minimize language-irrelevant variance attributable to "yes" vs. "no" decision processing. Fixation trials (24 for each run) were included as a baseline and required the participant to press the "yes" button when a fixation-cross at the center of the screen turned from red to blue. Perceptual trials (12 for each run) were included in each run to permit contrasts against perceptual processing for a related study but were not used in the present study. These perceptual trials comprised two sequences containing tones (AA), or tones followed by non-alphabetic characters (AV). These stimuli were presented as increasing, decreasing or steady in pitch (for auditory stimuli) or height (for visual stimuli). Participants were required to indicate *via* button press whether the sequences matched (e.g., two rising sequences) or mismatched (e.g., a rising sequence followed by a falling sequence). The timing for the fixation and perceptual trials were the same as for the lexical trials. Each run had a different pseudo-randomly interleaved ordering of lexical, fixation and perceptual trials. The intertrial intervals varied between 2,200 and 2,800 and optimized for an event-related design using OptSeq[2] to facilitate the modeling of overlapping hemodynamic responses. The lists were fixed across participants.

Functional MRI Data Acquisition

Participants were positioned in the MRI scanner with their head secured using foam pads and outfitted with an optical response box in the right hand. Visual stimuli were projected onto a rear-mounted screen viewed from a mirror attached to the inside of the head coil. Participants wore sound attenuating headphones to minimize the effects of the ambient scanner noise and deliver the auditory stimuli. Images were acquired using a 3.0 Tesla Siemens Trio scanner. A high resolution T1-weighted 3D structural image was first acquired for each subject (TR = 1,570 ms, TE = 3.36 ms, matrix size = 256 × 256, field of view = 240 mm, slice thickness = 1 mm, voxel size = 1 mm × 1 mm, number of slices = 160). Blood Oxygen Level Dependent (BOLD) functional images were acquired using a single-shot EPI (echo planar imaging) method, interleaved obliquely from bottom to top in a whole-brain acquisition (TE = 20 ms, flip angle = 80°, matrix size = 128 × 120, field of view = 220 × 206.25 mm, slice thickness = 3 mm (0.48 mm gap), number of slices = 32, TR = 2,000 ms, voxel size = 1.72 mm × 1.72 mm).

---

[1] http://elexicon.wustl.edu

[2] http://surfer.nmr.mgh.harvard.edu/optseq

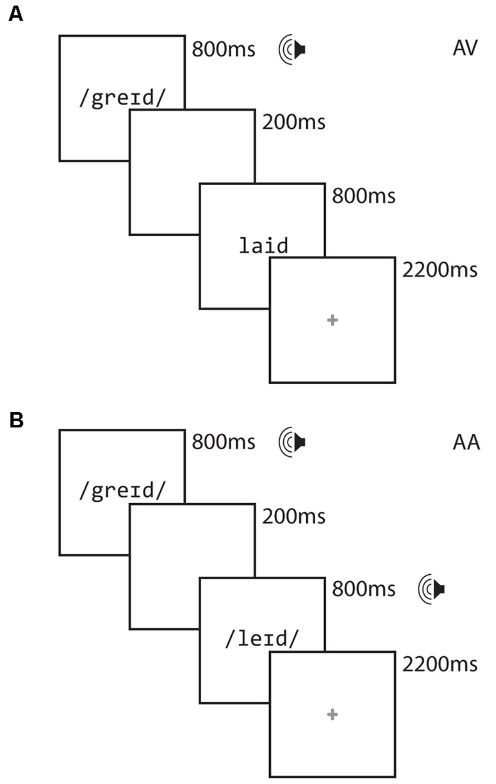


FIGURE 1 | Task diagram for the AV Crossmodal task **(A)** and AA Unimodal task **(B)**.

Functional MRI Data Preprocessing

fMRI data were analyzed using SPM8 (Statistical Parametric Mapping[3]). ArtRepair software[4] was used during image preprocessing to correct for participant movement by replacing outlier volumes associated with interpolated values from the two adjacent non-outlier scans. Outlier scans were defined as those for which a signal change of more than 1.5% from the mean, or movement of more than one voxel along any axis was detected. No more than 10% of the volumes from each run and no more than four consecutive volumes were interpolated in this way. A single attempt was made to reacquire runs requiring replacement of more than 10% of the volumes or more than four consecutive volumes. Slice time correction was applied to minimize timing-errors between slices. Functional images were co-registered with the anatomical image and normalized to the Montreal Neurological Institute (MNI) ICBM152 T1 template.

Design and Analysis

A standard general linear model (GLM) analysis estimated the neural response associated with each experimental condition at each voxel within the brain by convolving the vector of event onsets for each of the trial types (four lexical, one perceptual and one fixation) with SPMs canonical hemodynamic response function (HRF), and using the convolved waveforms to predict the observed BOLD waveforms in each voxel. The goodness of fit between convolved and observed waveforms was thus computed as a standardized regression coefficient (beta) for each of the six trial types in a GLM predicting voxel-wise BOLD activation. These betas are conventionally used as a measure of the responsiveness of the neural populations within each voxel to each of the trial types.

A first-level GLM analysis was performed for each participant and included t-tests statistically contrasting the responsiveness of each voxel to the lexical conditions and to the fixation baseline (LEX > NULL). A second-level random-effects analysis of the single-subject (LEX > NULL) contrast followed, collapsing across all participants to verify that the pattern of fMRI activations for our sample was in-line with those reported in previous studies.

Our previous study had found differences between TD and RD children with respect to audiovisual integration in STG, FG, IPL and STS (McNorgan et al., 2013). We note that the group-level GLM analyses will show significant task-related activations in other brain areas, however, we did not include these additional regions in our region of interest (ROI) analysis. We took this approach primarily for two reasons: first, not all regions have theoretical ties to reading development or developmental reading disorders, making it difficult to interpret results associated with these regions. Second, though there may be significant task activations in a region, it does not follow that these activations will correlate with reading skills. Thus, because we applied a Bonferroni correction for multiple comparisons to our ROI analyses, including these regions would increase the Type II error rate with a diminishing likelihood of gaining novel theoretical insight. For these reasons, restricted our analyses to those reasons for which we had *a priori* hypothesis.

ROIs were generated for each participant, allowing us to identify and characterize the neural activations in atlas-based definitions of these regions taken from the Wake Forest University PickAtlas, which was also used to help identify peak activations in the GLM analysis[5]. Because the PickAtlas provides gyral but not sulcal definitions, 4 mm dilations of the STG and MTG PickAtlas regions were intersected to generate an atlas-based definition of the STS, from which the posterior-most third was taken as the pSTS, as we have done in our previously published studies exploring this region. The anatomical extents of these atlas-based definitions are illustrated in **Figure 2**. Participant-specific ROIs were generated separately for the AA and AV modalities (i.e., two sets of masks) by identifying within each of these anatomical regions those subsets of voxels showing numerically greater activation for all lexical conditions than for the fixation condition within that modality—that is, no participant's ROIs included all the voxels included within these anatomical masks, but rather, these anatomical masks ensured that the functionally-defined ROIs for each individual were constrained to those anatomical regions prescribed by the hypotheses we were testing. An absolute statistical threshold was not applied because conventional statistical significance thresholds (e.g., $p < 0.05$)

---

[3] https://www.fil.ion.ucl.ac.uk/spm/

[4] http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html

[5] http://fmri.wfubmc.edu/software/pickatlas

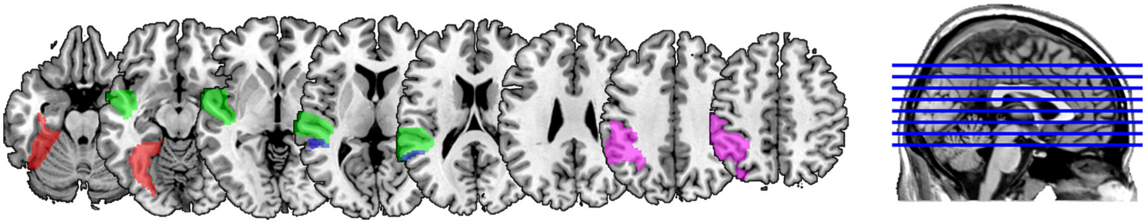


FIGURE 2 | Anatomical extents of the atlas-based anatomical definitions of the masks used to constrain region of interest (ROI) definitions for fusiform gyrus (FG; red), superior temporal gyrus (STG; green), superior temporal sulcus (STS; blue) and inferior parietal lobule (IPL; magenta). Voxels falling within each of these regions that showed greater activity for lexical trials vs. baseline for a participant were included in that participant's ROI for that anatomical label.

failed to select voxels in all regions for all participants, who notably come from a population for which reduced activation is commonly found among these regions (Richlan et al., 2009). These masks identified for each child voxels within these anatomical regions showing heightened activity under either unimodal AA and crossmodal AV task conditions, respectively. Importantly, because the same voxel selection criterion was used for all conditions and all participants, the ROI masks were bias-free.

We calculated mean signal strength across all rhyming lexical conditions and for the fixation baseline condition in each of the four regions separately for each task modality. Each calculation used the ROI mask for the corresponding task modality. Thus, for example, when computing the mean signal strength within the AA task data, the mean value for the FG was calculated overall FG voxels showing greater than baseline lexical activation in the AA condition, whereas, for AV task modality, this value was calculated overall FG voxels showing greater than baseline lexical activation in the AV condition. By calculating signal strength in this way, we avoided misleading comparisons between modalities that might arise from the assumption of similar spatial distributions of positive activations. Such an assumption could lead us to omit many relevant voxels or include many irrelevant voxels. Instead, we focused only and exactly on those voxels with any degree of positive association with the task for each modality condition.

Our ROI analysis submitted mean signal strength for the AA and AV task modality conditions and baseline signal strength and performance IQ to a hierarchical multiple regression with either raw Blending scores or raw Elision scores as the dependent measure of phonemic awareness. Because McNorgan et al. (2013) previously found no relation between audio-visual processing and Elision scores among these regions, our focus was in determining whether skill in the synthetic aspect of phonemic awareness might instead be predicted by unimodal processing, after accounting for variance predicted by nuisance regressors and by audio-visual processing. The sequence of regression steps forced age (in months), baseline signal and performance IQ in the first block as nuisance regressors, mean AV signal strength in the second block, and mean AA signal in the final block, predicting phonemic awareness as a function of the neural activity associated with both unimodal and crossmodal language processing. This approach conservatively controls for baseline signal strength, participant age, and performance IQ nuisance regressors but maximized sensitivity to any predictive ability of neural activity during the AV task. However, given the previously reported null effects for audiovisual processing in this population, we focus on whether unimodal processing during the AA task significantly predicts phonemic awareness after controlling for our nuisance predictors and audio-visual processing. The regression analyses were performed for each ROI, controlling for multiple comparisons, allowing us to determine whether task-related activity predicted synthetic phonemic awareness in each region. Because McNorgan et al. (2011) showed that dyslexic children demonstrate an atypical pattern of audiovisual integration-related processing, even under crossmodal conditions that should promote audiovisual integration, we predicted that phonemic awareness would be related to processing in the unimodal AA task condition, but not the crossmodal AV task condition, and only in regions associated with unimodal processing.

RESULTS

GLM Analysis

Figures 3, **4** illustrate significant clusters in the group-level GLM analysis using an uncorrected voxel-wise significance threshold of $p < 0.001$ and a cluster-level family-wise error rate of $p < 0.05$ (i.e., clusters of the obtained size are 5% likely to occur by chance under Gaussian random field theory). Focusing on the modality conditions in isolation (**Figure 3**), the analysis found both the unimodal AA and crossmodal AV task conditions were associated with activations significantly above baseline in a network of regions implicated in phonological processing (bilateral BA 21/22; STG) and visual/orthographic processing (BA 17/18/37; Cuneus, extending into FG and left calcarine fissure). Additionally, the crossmodal condition was associated with significant clusters in left IFG (BA 44/45; Broca's area) and left precentral gyrus. Direct contrasts between the two task modality conditions (**Figure 4**) found the unimodal task was associated with significantly greater activation only in bilateral STG, whereas the crossmodal task was associated with significantly greater activation in bilateral occipitotemporal and left inferior frontal regions. The coordinates of peak maxima for these clusters are presented

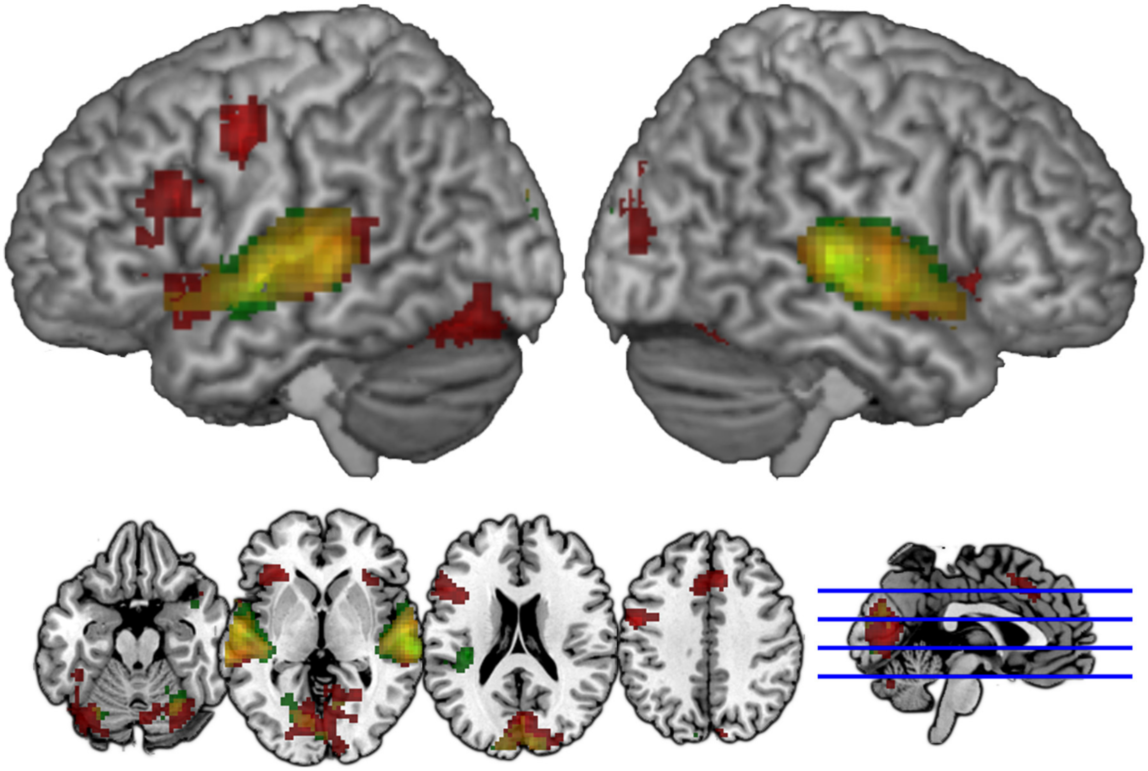


FIGURE 3 | Spatial distribution of voxels demonstrating group-level lexicality effects (contrast of Lexical vs. Fixation trials) in the AV Crossmodal task (red) and AA Unimodal task (green). Overlapping modality effects appear in yellow. Clusters are extent-corrected at an FWE significance level of $p < 0.05$, with an uncorrected voxel-wise $p < 0.001$.

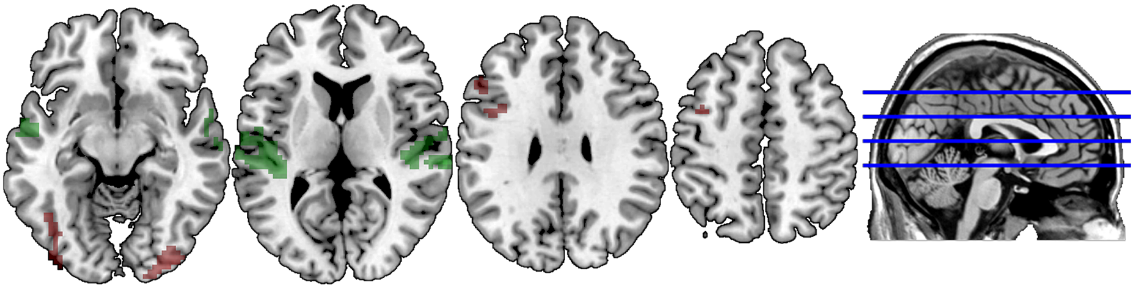


FIGURE 4 | Spatial distribution of voxels demonstrating significantly greater lexicality effects in the Unimodal task (green) vs. the Crossmodal task, and demonstrating significantly greater lexicality effects in the Crossmodal task (red) vs. the Unimodal task. Clusters are extent-corrected at an FWE significance level of $p < 0.05$, with an uncorrected voxel-wise $p < 0.001$.

in **Table 2**. These overall results indicate that participants were engaging networks of regions commonly associated with phonological and orthographic processing of language, and did so under both task modality conditions. The task modality contrasts reflect the relative orthographic and phonological demands associated with each task modality: The unimodal task modality placed greater demand on bilateral primary and associative auditory processing regions, whereas the crossmodal task placed greater demand on bilateral visual processing regions and the left inferior frontal gyrus, which has been argued to play a critical role in visual word recognition (Cornelissen et al., 2009).

ROI Analysis

The exclusionary criteria for this study selected children clinically diagnosed with RD and no other cognitive or behavioral impairment. Consequently, though performance IQ and baseline fixation activity were included as nuisance regressors in the analyses that follow, it is unsurprising that neither were significant predictors in any of the regression

TABLE 2 | Table of coordinates of peak activations for significant extent-corrected clusters of activation.

Region	Size	pFWE	Max Z	*x*	*y*	*z*
Unimodal						
R. Superior/Middle Temporal Gyrus (BA 22, 21)	861	<0.001	6.58	63	−27	3
L. Superior/Middle Temporal Gyrus (BA 22, 21)	1,050	<0.001	6.09	−57	−9	−3
R. Cuneus (BA 18)	515	<0.001	4.66	9	−90	24
R. Cerebellum	189	<0.001	4.51	27	−63	−24
Crossmodal						
R. Superior/Middle Temporal Gyrus (BA 22, 21)	703	<0.001	5.92	60	−12	3
L. Superior/Middle Temporal Gyrus (BA 22, 21)	1,009	<0.001	5.78	−54	−27	3
L. Precentral Gyrus (BA 6)	139	<0.001	5.15	−48	−3	42
R. Calcarine Sulcus (BA 17)	1,853	<0.001	5.05	3	−72	15
R. Middle Cingulum (BA 32)	239	<0.001	4.99	12	27	36
L. Inferior Frontal Gyrus (BA 44/45)	163	<0.001	4.85	−51	18	24
R. Insula (BA 13)	68	0.012	4.63	39	21	−9
Crossmodal > Unimodal						
L. Fusiform Gyrus (BA 37)	99	0.002	5.41	−42	−66	−15
L. Inferior Frontal Gyrus (BA 45)	147	<0.001	5.08	−45	18	24
R. Inferior Occipital Gyrus (BA 18)	91	0.002	4.15	33	−93	−9
Unimodal > Crossmodal						
R. Superior Temporal Gyrus (BA 22, 21)	349	<0.001	5.55	57	−18	0
L. Superior Temporal Gyrus (BA 22, 21)	431	<0.001	4.71	−57	−9	−3

Note: L, left; R, right; BA, Brodmann Area (approx.); FDR, FDR-corrected significance level; Max, maximum. Size is measured in voxels. Coordinates reflect standard MNI space.

analyses that follow, nor was age. Scatterplot diagrams for the regressions predicting Blending and Elision scores are presented in **Figures 5**, **6**, respectively. All significance levels are reported using a Šidák family-wise error rate correction for multiple comparisons.

Superior Temporal Gyrus

The hierarchical linear regression found mean unimodal lexical task-related activity in the STG significantly improved prediction of raw Blending scores after accounting for all other regressor variables ($F_{(1,12)} = 7.33$, $p = 0.038$. In the final model, only unimodal activity was a significant predictor of raw Blending scores ($t_{(12)} = 2.71$, $p = 0.019$, partial $r_{(16)} = 0.48$, $f^2 = 0.96$). Mean crossmodal lexical task-related activity in the STG was not a significant predictor of Blending scores ($t_{(12)} = 0.641$, $p = 0.54$). This pattern of results suggests that neural processing in STG is related to phoneme synthesis in children with RD under unimodal, but not crossmodal task conditions. The hierarchical linear regression found mean unimodal lexical task-related activity in the STG did not significantly improve the prediction of raw Elision scores after accounting for all other regressor variables ($F_{(1,12)} = 2.887$, $p = 0.52$. In the final model, neither mean unimodal activity ($t_{(12)} = 1.699$, $p = 0.115$) nor crossmodal activity in the STG ($t_{(12)} = -0.279$, $p = 0.79$) were significant predictors of raw Elision scores. This pattern of results suggests that neural processing in STG is unrelated to phonemic analysis ability in children with RD under either unimodal or crossmodal task conditions.

Fusiform Gyrus

The hierarchical linear regression found mean unimodal lexical task-related activity in the FG did not significantly improve the prediction of raw Blending scores after accounting for all other regressor variables ($F_{(1,12)} = 0.193$, $p = 0.999$. In the final model, neither mean unimodal activity ($t_{(12)} = -0.439$, $p = 0.67$) nor crossmodal activity in the FG ($t_{(12)} = -0.497$, $p = 0.63$) were significant predictors of raw Blending scores. This pattern of results suggests that neural processing in FG is unrelated to phoneme synthesis in children with RD under either unimodal or crossmodal task conditions. The hierarchical linear regression found mean unimodal lexical task-related activity in the FG marginally improved prediction of raw Elision scores after accounting for all other regressor variables ($F_{(1,12)} = 7.51$, but because we anticipated a null effect, this improvement was not significant after the familywise error rate correction was applied, $p = 0.10$. In the final model, mean unimodal activity ($t_{(12)} = 2.739$, $p = 0.018$, partial $r_{(16)} = 0.59$, $f^2 = 1.44$) was a significant predictor of raw Elision scores, but crossmodal activity in the FG ($t_{(12)} = 1.015$, $p = 0.22$) was not. This pattern of results suggests that neural processing in FG may be weakly related to phoneme analysis ability in children with RD under unimodal but not crossmodal task conditions.

Superior Temporal Sulcus

The hierarchical linear regression found mean unimodal lexical task-related activity in the STS did not significantly improve the prediction of raw Blending scores after accounting for all other regressor variables ($F_{(1,12)} = 2.109$, $p = 0.32$. In the final model, neither mean unimodal activity ($t_{(12)} = 1.452$, $p = 0.17$) nor crossmodal activity in the STS ($t_{(12)} = -0.889$, $p = 0.39$) were significant predictors of raw Blending scores. This pattern of results suggests that neural processing in STS is unrelated to phoneme synthesis in children with RD under either unimodal or crossmodal task conditions. The hierarchical linear regression found mean unimodal lexical task-related activity in the STS did not significantly improve the prediction of raw Elision scores after accounting for all other regressor variables ($F_{(1,12)} = 0.744$, $p = 0.96$. In the final model, neither mean unimodal activity ($t_{(12)} = 0.863$, $p = 0.41$) nor crossmodal activity in the STS ($t_{(12)} = -0.339$, $p = 0.74$) were significant predictors of raw

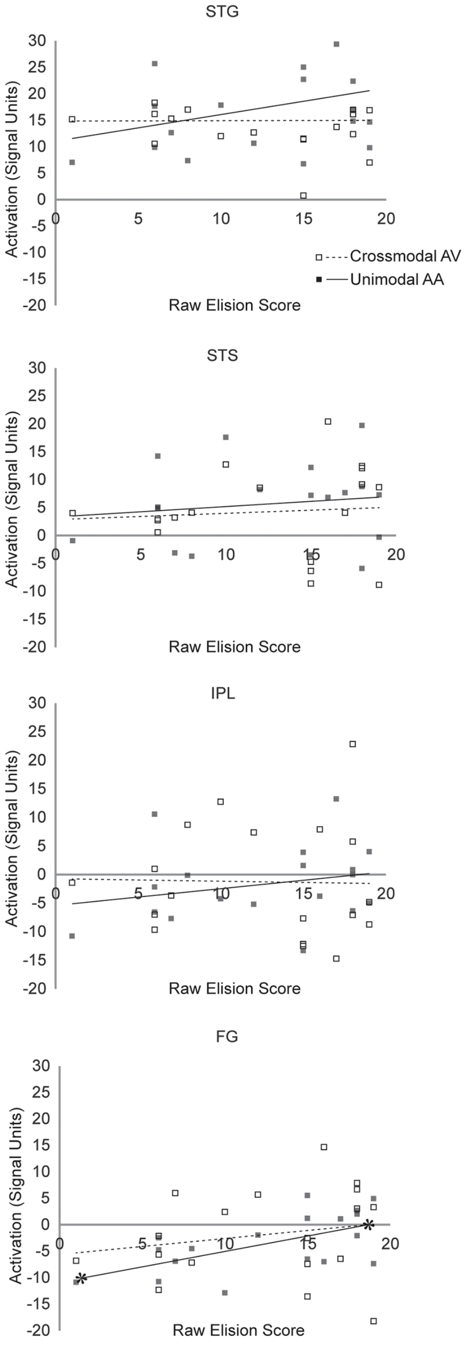


FIGURE 5 | Scatterplot diagram of ROI activations as a function of Blending scores. Significant regression lines are capped with asterisks.

Elision scores. This pattern of results suggests that neural processing in STS is unrelated to phoneme analysis in children with RD under either unimodal or crossmodal task conditions.

Inferior Parietal Lobule

The hierarchical linear regression found mean unimodal lexical task-related activity in the IPL did not significantly improve the prediction of raw Blending scores after accounting for all other regressor variables ($F_{(1,12)} = 0.09$, $p = 0.99$. In the final model, neither mean unimodal activity ($t_{(12)} = -0.304$, $p = 0.77$) nor crossmodal activity in the IPL ($t_{(12)} = -0.732$, $p = 0.48$) were significant predictors of raw Blending scores. This pattern of results suggests that neural processing in IPL is unrelated to phoneme synthesis in children with RD under either unimodal or crossmodal task conditions.

The hierarchical linear regression found mean unimodal lexical task-related activity in the IPL did not significantly improve the prediction of raw Elision scores after accounting for all other regressor variables ($F_{(1,12)} = 0.744$, $p = 0.93$. In the final model, neither mean unimodal activity ($t_{(12)} = 2.73$, $p = 0.36$) nor crossmodal activity in the IPL ($t_{(12)} = 1.015$, $p = 0.58$) were significant predictors of raw Elision scores. This pattern of results suggests that neural processing in IPL is unrelated to phoneme analysis in children with RD under either unimodal or crossmodal task conditions.

DISCUSSION

The purpose of this study was to further investigate how online rhyme judgment under unimodal and crossmodal presentation conditions predicts a continuum of phonemic awareness ability in children with RD. We examined the relationship between measures of phoneme synthesis (Blending) and analysis (Elision) with a rhyme judgment task presented in unimodal auditory-only (AA) and crossmodal audiovisual (AV) presentations in a left hemisphere sub-network of reading regions including STG, pSTS, IPL, and FG. We predicted that children with RD would demonstrate a reliance on unimodal processing in unisensory regions, but would not show a similar reliance on crossmodal processing in known multisensory processing sites. Based on previous findings reported by McNorgan et al. (2013), we anticipated that Elision would be unrelated to neural processing throughout this network. The results indicate that, for children with RD, phoneme synthesis *via* blending phonemes into whole word representations is related to STG activity during unimodal rhyme judgment, and that this is a large effect that is likely to replicate. The results also suggested that phoneme analysis *via* the Elision task may be similarly modulated by unisensory regions in the FG during unimodal rhyme judgment, and, though this was also found to be a large effect, we interpret this finding cautiously. This pattern suggests that better phonemic awareness in children with RD is associated with unimodal phonological processing, and implies a reliance on unisensory rather than multisensory brain regions to resolve these phonemic awareness tasks.

These results extend previous literature regarding the relationship between crossmodal rhyme judgment and phonological awareness. McNorgan et al. (2013) found the reciprocal pattern of findings to those we describe here: TD, but not RD readers, demonstrated a significant relationship between crossmodal AV rhyming and phoneme analysis skills in multisensory brain areas. Those results demonstrated a disconnect between phoneme analysis in an Elision task and crossmodal congruency in RD readers. However, the null effect previously described emphasized brain processes present in TD readers that appeared to be absent in RD

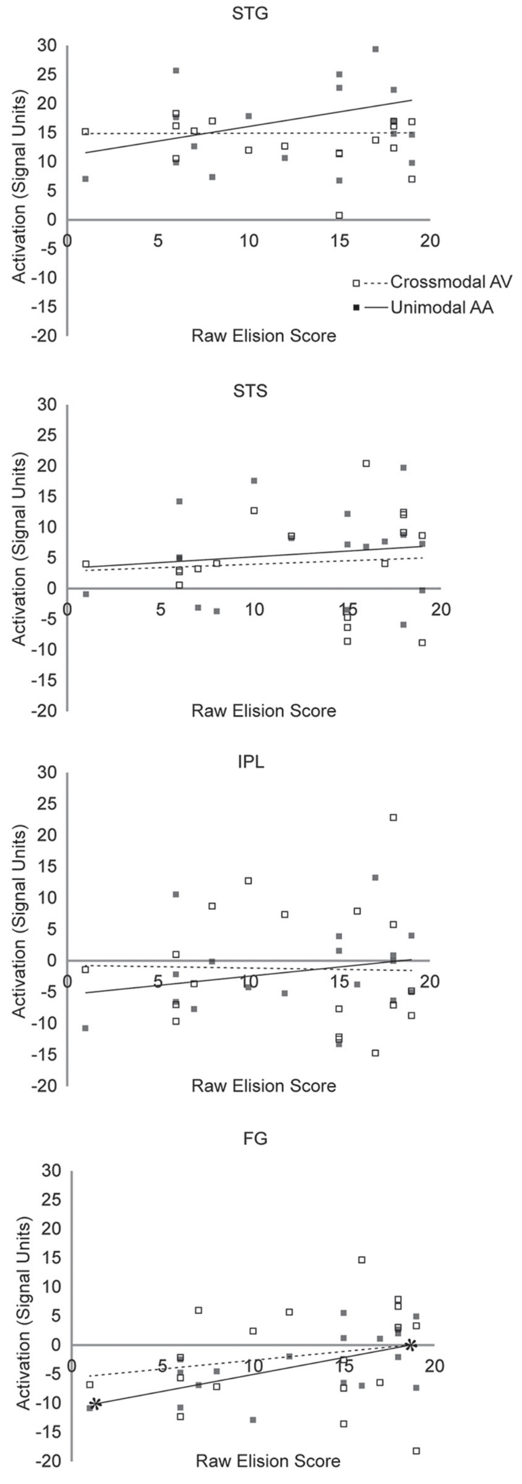


FIGURE 6 | Scatterplot diagram of ROI activations as a function of Elision scores. Significant regression lines are capped with asterisks.

readers (McNorgan et al., 2013), providing only indirect insight into phoneme analysis in this population. The present study provides further insight into the dynamics of the neural processes in which lower-skill readers *do* apply phonemic awareness skills to online lexical processing. Specifically, lower-skill readers with higher phonemic awareness appear to engage unisensory processing regions to perform auditory rhyme judgments.

Phonemic Awareness in Unisensory Brain Areas

Whereas reading is a quintessentially multisensory task—printed words are mapped to phonological representations, which, in turn, refine the orthographic system—it is important to bear in mind that the rhyming judgment task used here is phonologically-based. Synthesis of individual phoneme segments into a whole word relies heavily on phonological representation as well as phonological memory and retrieval in the STG (Simos et al., 2000; Temple et al., 2001; Turkeltaub et al., 2003). Phoneme synthesis is an early pre-literate skill that is predictive of reading ability (Perfetti et al., 1987; Ouellette and Haley, 2013). A straightforward interpretation of this result is that children with RD readily engage the phonological system involved in phoneme synthesis when making a phonological (rhyme) judgment. The features over which phonological similarity is evaluated may include not only basic acoustic information but also information likely to be tapped by phoneme synthesis (e.g., the sequence in which phonemes are combined). While these processes would not necessarily be facilitated by orthographic representation in the auditory-only condition in our task, converging evidence in TD readers shows that with increased reading skills the STG plays a greater role in visual letter processing (Blau et al., 2010). Fluent readers with stronger phonemic awareness skills show greater co-activation for print and speech processing in the STG (Frost et al., 2009; Preston et al., 2016). The present study adds to the body of evidence suggesting that RD is associated with less influence of visual input on phonological processing because processing under the AV presentation condition in STG—or any other ROI—was unrelated to any measure of phonological awareness.

Though the hierarchical regression failed to show that neural activity under the AA condition significantly improved prediction of phoneme analysis over AV neural activity and other nuisance regressors, the final model nonetheless indicated a significant relationship between phoneme analysis in the FG with unimodal lexical processing in the AA condition. We thus interpret this relationship cautiously, and in light of the large body of literature shows that RD is associated with under-activation of this region (for a review see McCandliss and Noble, 2003). Our finding does not contradict this research as this region was not activated above baseline for the auditory-only condition for the overall sample. Rather, this activation bore a linear relationship to the child's phonemic awareness. Because we did not find a similar relationship in the AV condition, this indicates that orthographic processing in this region is driven by mechanisms other than phonemic awareness when visual orthographic input is available, and thus that phonemic awareness does not mediate the resolution of written words in children with RD. This is not to suggest that children with RD do not engage the orthographic system during reading; rather it would be consistent with a model of RD in which the orthographic system does not provide ongoing support in the resolution of phonology (McNorgan et al., 2014). This may be the optimal strategy if audiovisual integration processes fail to generate useful information from the orthographic representation.

We take the well-supported position that developmental reading difficulties arise in large part from an inability to accurately and quickly map between phonological and orthographic representations (Booth et al., 2004, 2007; Cao et al., 2006; Bitan et al., 2007). Multisensory interactions provide useful constraints on lexical activations in either modality to the extent that they help disambiguate multiple competing representations (Seidenberg and McClelland, 1989; Harm and Seidenberg, 1999). Indeed, Price and Devlin's (2011) Interactive Account argues that the specialization of FG for orthographic processing is a consequence of internally-driven (i.e., top-down) phonological input facilitating the perceptually-driven (i.e., bottom-up) visual object processing system. This facilitation hinges on effective crossmodal integration. The introduction of additional disambiguating information helps reduce uncertainty and allows a clear winning lexical representation to emerge. It follows that ineffective crossmodal integration may provide no useful information, or even misleading information.

Audiovisual Integration Underlies Mapping Between Orthography and Phonology

The results described above, along with the underactivation of the IPL and pSTS, indicate that the children with RD in our sample may not have effectively extracted the statistical regularities in the mapping between orthography and phonology. Children with RD may utilize more direct access to whole word representations for auditory rhyme judgment, rather than operating at a smaller grain size that would refine by higher phonemic analysis skills and more specialized orthographic processing (Ziegler and Goswami, 2005). As anticipated, we did not find a relationship between phoneme analysis and neural activity in multisensory integration sites, such as the pSTS and IPL. If multisensory integration within these regions is critical for mapping between phonological and orthographic representations, this pattern may explain the failure to find a relationship between phonemic awareness and audiovisual processing for our RD readers: continuity between the phonological, multisensory, and orthographic systems would imply that phonemic awareness is related to processing across all three systems. Because phonemic awareness is unrelated to processing in multisensory brain regions and in the crossmodal conditions, this suggests that the IPL or pSTS do not contribute towards mapping between orthographic and phonological representations in children with RD. Thus, though phonemic awareness may influence phonological and orthographic processing in children with RD, it does so without the coordination that audiovisual integration processes afford.

The results of the current study provide further support for the hypothesis that crossmodal integration between letters and speech sounds is impaired in children with RD. A large body of literature has shown a failure to integrate letters and speech sounds as a causal factor in dyslexia. Failure to integrate individual phonemes with graphemes in transparent orthographies such as Dutch and German has been documented in both event-related potential (ERP) and fMRI studies of children during literacy acquisition (van Atteveldt et al., 2009; Blau et al., 2010; Blomert, 2011; Bakos et al., 2017). However, a recent ERP study of English-speaking children challenged the letter-sound integration hypothesis (Nash et al., 2017). The authors found only mild deficits for letter-sound integration in RD children compared to age and reading skill matched children. One explanation of the apparent inconsistency between these findings and proponents of the letter-sound integration hypothesis is that dyslexia manifests differently in shallow and deep orthographies. Dyslexia in shallow orthographies, like Dutch, may be characterized by slow, effortful serial processing of letters, while in deep orthographies, such as English, the slow speed and effort may be at a larger grain size, such as the rhymes tested in our phonological task. The suggestion that orthographic depth likely interacts with RD is supported by a recent neuroimaging study that found distinct areas of under-activation in shallow and deep orthographies (Martin et al., 2016). However, in both types of orthographic systems, low skilled readers under-activate the occipitotemporal cortex. This under-activation of unimodal visual areas implies a reliance on access to unimodal representations, which may be degraded in RD. The results of the current study indicate that phonemic awareness, particularly phonemic analysis, is not active in the binding of orthographic and phonological representations for children with RD. The children with RD in the current study were near adolescent, spanning a range of reading ability, with some near-average performance on standardized tasks. Thus, the neurobiological profile outlined in this study signifies a persistence of deficits in mapping between modalities even after several years of reading instruction at school. This indicates that similar to visual word recognition, the neural processing engaged in phoneme synthesis and analysis relies on alternative mechanisms in children with RD.

Limitations and Future Directions

The results of the current study suggest that both phoneme synthesis and analysis in children with RD rely on unisensory brain areas and unimodal processing. However, phonemic awareness as measured in the current study may be mediated by other factors such as attention, working memory, and overall language ability. The design of the between-subjects study was a within-subjects examination of how phonemic awareness skill is related to crossmodal processing. Future between-subjects experiments may utilize a broader range of phonemic awareness tasks (e.g., deletion, segmentation, letter rhyming) in addition to functional skills like reading fluency and comprehension. Experiments such as these, using appropriately matched groups, would support explorations of how phonemic awareness might differentially support reading development in RD and TD populations. Similarly, as the diagnostic labels associated with

RD imply a heterogeneous range of deficits related to phonology, semantics, print processing, and general linguistic ability, larger studies may further examine individual differences in crossmodal lexical processing.

We constrained our ROIs to those regions that are proposed to have a specific role in the processing of orthography, phonology, or the integration of these elements, aligned with van Atteveldt et al.'s (2004) model of the left posterior integration network. Future studies may explore the relationship between right hemisphere structures and crossmodal processing of orthography and phonology in RD. Hemispheric differences between TD and RD readers are apparent, and may indicate compensatory mechanisms in RD in the right hemisphere (Pugh et al., 2001; Démonet et al., 2004; Shaywitz and Shaywitz, 2005). For phonological judgment tasks, children with RD demonstrate enhanced activity in the right compared to left inferior temporal gyrus (Corina et al., 2001). Activity in the right anterior insula and right STS are enhanced in adults with dyslexia compared to those without for audiovisual and visual lexical judgment, indicating that the right hemisphere recruitment of homologous structures occurs during crossmodal lexical processing (Kast et al., 2011). Examination of the right hemisphere recruitment related to phonological and crossmodal processing may further inform the understanding of how RD readers apply orthographic and phonological representations to lexical processing. A test of whether this compensatory recruitment varies between word and pseudoword trials may provide insight into whether these activations are driven by visual or semantic processing (or both) as argued by Pugh et al. (2001).

SUMMARY

We explored the relationship between phonemic awareness and modality presentation in children with RD along a continuum of reading ability. Previous fMRI studies have found that in TD readers, phonemic awareness skill is associated with crossmodal integration of phonology and orthography (Frost et al., 2009; McNorgan et al., 2013). For RD readers, we did not find any association between brain activation in crossmodal (AV) tasks and phonemic awareness. However, we did find significant brain-behavior correlations in the STG for the phonemic awareness measure of Blending with unimodal auditory-only presentation and in the FG with Elision with the unimodal auditory-only presentation. These significant brain-behavior correlations were found in unisensory areas implicated in the processing of orthography (FG) and phonology (STG). Using a hypothesis-driven, ROI-based approach, we did not find any significant correlations for pSTS and IPL, areas implicated in crossmodal integration across a number of domains, including language processing. Future studies may further examine the functional connectivity within this reading network to further elucidate how connectivity between these crossmodal regions are predictive of phonemic awareness in both high and low skilled readers.

Despite some of our RD participants having near-typical performance on standardized measures of phonemic awareness, as a group, our sample does not show a relationship between phonemic awareness and crossmodal integration in multisensory regions as found in TD children (McNorgan et al., 2013). Rather, the children with RD show a relationship between phonemic awareness and unimodal auditory processing in unisensory STG. This indicates that phonemic awareness remains related to phonological processing, but is not related to the integration of orthographic and phonological representations in RD readers, even after approximately 5–9 years of reading instruction at school. The educational implications of these results indicate that phonemic awareness skills, particularly performance in phoneme analysis is not reflective of advancing literacy skills in RD, and may rather be mediated by alternative strategies (Shaywitz et al., 2003). As such, educators and interventionists need to be careful in the interpretation of how phonemic awareness constructs relate to reading ability when designing instruction.

DATA AVAILABILITY STATEMENT

Data are available at Open Neuro (https://openneuro.org/datasets/ds001894/versions/1.3.0).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board at Northwestern University, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board at Northwestern University.

AUTHOR CONTRIBUTIONS

MR, CM, and JB conceptualized the study. MR and CM were involved in data collection. CM and EG carried out the data analysis and created the figures. MR, CM, EG, and JB contributed to the drafting and editing of the final manuscript.

FUNDING

This research was supported by grants from the National Institute of Child Health and Human Development (HD042049) to JB.

ACKNOWLEDGMENTS

We would like to thank the many research assistants and colleagues who assisted in the data collection throughout the study.

REFERENCES

Anthony, J. L., and Lonigan, C. J. (2004). The nature of phonological awareness: converging evidence from four studies of preschool and early grade school children. *J. Educ. Psychol.* 96, 43–55. doi: 10.1037/0022-0663.96.1.43

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., and Burgess, S. R. (2003). Phonological sensitivity: a quasi-parallel progression of word structure units and cognitive operations. *Read. Res. Q.* 38, 470–487. doi: 10.1598/rrq.38.4.3

Bakos, S., Landerl, K., Bartling, J., Schulte-Körne, G., and Moll, K. (2017). Deficits in letter-speech sound associations but intact visual conflict processing in dyslexia: results froma Novel ERP-paradigm. *Front. Hum. Neurosci.* 11:116. doi: 10.3389/fnhum.2017.00116

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4

Bitan, T., Burman, D. D., Chou, T. L., Lu, D., Cone, N. E., Cao, F., et al. (2007). The interaction between orthographic and phonological information in children: an fMRI study. *Hum. Brain Mapp.* 28, 880–891. doi: 10.1002/hbm.20313

Blau, V., Reithler, J., van Atteveldt, N. M., Seitz, J., Gerretsen, P., Goebel, R., et al. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133, 868–879. doi: 10.1093/brain/awp308

Blau, V., van Atteveldt, N. M., Ekkebus, M., Goebel, R., and Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr. Biol.* 19, 503–508. doi: 10.1016/j.cub.2009.01.065

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage* 57, 695–703. doi: 10.1016/j.neuroimage.2010.11.003

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002). Modality independence of word comprehension. *Hum. Brain Mapp.* 16, 251–261. doi: 10.1002/hbm.10054

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2004). Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* 16, 1234–1249. doi: 10.1162/0898929041920496

Booth, J. R., Cho, S., Burman, D. D., and Bitan, T. (2007). Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. *Dev. Sci.* 10, 441–451. doi: 10.1111/j.1467-7687.2007.00598.x

Burgess, S. R., and Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: evidence from a preschool sample. *J. Exp. Child Psychol.* 70, 117–141. doi: 10.1006/jecp.1998.2450

Bus, A. G., and van IJzendoorn, M. H. (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *J. Educ. Psychol.* 91, 403–414. doi: 10.1037/0022-0663.91.3.403

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi: 10.1093/cercor/11.12.1110

Cao, F., Bitan, T., Chou, T. L., Burman, D. D., and Booth, J. R. (2006). Deficient orthographic and phonological representations in children with dyslexia revealed by brain activation patterns. *J. Child Psychol. Psychiatry* 47, 1041–1050. doi: 10.1111/j.1469-7610.2006.01684.x

Castles, A., Holmes, V., Neath, J., and Kinoshita, S. (2003). How does orthographic knowledge influence performance on phonological awareness tasks? *Q. J. Exp. Psychol. A* 56, 445–467. doi: 10.1080/02724980244000486

Castles, A., Wilson, K., and Coltheart, M. (2011). Early orthographic influences on phonemic awareness tasks: Evidence from a preschool training study. *J. Exp. Child Psychol.* 108, 203–210. doi: 10.1016/j.jecp.2010.07.006

Corina, D. P., Richards, T. L., and Serafini, S. (2001). fMRI auditory language differences between dyslexic and able reading children. *Neuroreport* 12, 1195–1201. doi: 10.1097/00001756-200105080-00029

Cornelissen, P. L., Kringelback, M. L., Ellis, A. W., Whitney, C., Holliday, I. E., and Hansen, P. C. (2009). Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from megnetoencephalography (MEG). *PLoS One* 4:e5359. doi: 10.1371/journal.pone.0005359

Dehaene, S., and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262. doi: 10.1016/j.tics.2011.04.003

Démonet, J. F., Taylor, M. J., and Chaix, Y. (2004). Developmental dyslexia. *Lancet* 363, 1451–1460. doi: 10.1016/S0140-6736(04)16106-0

Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., and Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Res.* 1356, 73–84. doi: 10.1016/j.brainres.2010.07.097

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., and Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: evidence from the national reading panel's meta-analysis. *Read. Res. Q.* 36, 250–287. doi: 10.1598/rrq.36.3.2

Ehri, L. C., and Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Appl. Psycholinguist.* 1, 371–385. doi: 10.1017/s0142716400009802

Fraga González, G., Žarić, G., Tijms, J., Bonte, M., and van der Molen, M. W. (2017). Contributions of letter-speech sound learning and visual print tuning to reading improvement: evidence from brain potential and dyslexia training studies. *Brain Sci.* 7:E10. doi: 10.3390/brainsci7010010

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001

Frost, S. J., Landi, N., Mencl, W. E., Sandak, R., Fulbright, R. K., Tejada, E. T., et al. (2009). Phonological awareness predicts activation patterns for print and speech. *Ann. Dyslexia* 59, 78–97. doi: 10.1007/s11881-009-0024-y

Froyen, D., Van Atteveldt, N. M., Bonte, M., and Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci. Lett.* 430, 23–28. doi: 10.1016/j.neulet.2007.10.014

Gullick, M. M., and Booth, J. R. (2014). Individual differences in crossmodal brain activity predict arcuate fasciculus connectivity in developing readers. *J. Cogn. Neurosci.* 26, 1331–1346. doi: 10.1162/jocn_a_00581

Harm, M. W., and Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychol. Rev.* 106, 491–528. doi: 10.1037/0033-295x.106.3.491

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/s1364-6613(00)01463-7

Humphries, C., Sabri, M., Lewis, K., and Liebenthal, E. (2014). Hierarchical organization of speech perception in human auditory cortex. *Front. Neurosci.* 8:406. doi: 10.3389/fnins.2014.00406

Kast, M., Bezzola, L., Jäncke, L., and Meyer, M. (2011). Multi-and unisensory decoding of words and nonwords result in differential brain responses in dyslexic and nondyslexic adults. *Brain Lang.* 119, 136–148. doi: 10.1016/j.bandl.2011.04.002

Kroese, J. M., Hynd, G. W., Knight, D. F., Hiemenz, J. R., and Hall, J. (2000). Clinical appraisal of spelling ability and its relationship to phonemic awareness (blending, segmenting, elision, and reversal), phonological memory and reading in reading disabled, ADHD, and normal children. *Read. Writ.* 13, 105–131. doi: 10.1023/A:1008042109851

Kronschnabel, J., Brem, S., Maurer, U., and Brandeis, D. (2014). The level of audiovisual print-speech integration deficits in dyslexia. *Neuropsychologia* 62, 245–261. doi: 10.1016/j.neuropsychologia.2014.07.024

Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 44, 978–990. doi: 10.3758/s13428-012-0210-4

Lonigan, C. J., Anthony, J. L., Phillips, B. M., Purpura, D. J., Wilson, S. B., and McQueen, J. D. (2009). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *J. Educ. Psychol.* 101, 345–358. doi: 10.1037/a0013837

Mann, V., and Wimmer, H. (2002). Phoneme awareness and pathways into literacy: a comparison of german and american children. *Read. Writ.* 15, 653–682. doi: 10.1023/A:1020984704781

Martin, A., Kronbichler, M., and Richlan, F. (2016). Dyslexic brain activation abnormalities in deep and shallow orthographies: a meta-analysis of

28 functional neuroimaging studies. *Hum. Brain Mapp.* 37, 2676–2699. doi: 10.1002/hbm.23202

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299. doi: 10.1016/s1364-6613(03)00134-7

McCandliss, B. D., and Noble, K. G. (2003). The development of reading impairment: a cognitive neuroscience model. *Ment. Retard. Dev. Disabil. Res. Rev.* 9, 196–205. doi: 10.1002/mrdd.10080

McNorgan, C., Alvarez, A., Bhullar, A., Gayda, J., and Booth, J. R. (2011). Prediction of reading skill several years later depends on age and brain region: implications for developmental models of reading. *J. Neurosci.* 31, 9641–9648. doi: 10.1523/JNEUROSCI.0334-11.2011

McNorgan, C., Awati, N., Desroches, A. S., and Booth, J. R. (2014). Multimodal lexical processing in auditory cortex is literacy skill dependent. *Cereb. Cortex* 24, 2464–2475. doi: 10.1093/cercor/bht100

McNorgan, C., Randazzo-Wagner, M., and Booth, J. R. (2013). Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Front. Hum. Neurosci.* 7:388. doi: 10.3389/fnhum.2013.00388

Melby-Lervåg, M., Lyster, S.-A. H., and Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138, 322–352. doi: 10.1037/a0026744

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., et al. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learn. Instr.* 29, 65–77. doi: 10.1016/j.learninstruc.2013.09.003

Nash, H. M., Gooch, D., Hulme, C., Mahajan, Y., McArthur, G., Steinmetzger, K., et al. (2017). Are the literacy difficulties that characterize developmental dyslexia associated with a failure to integrate letters and speech sounds? *Dev. Sci.* 20:e12423. doi: 10.1111/desc.12423

Ouellette, G. P., and Haley, A. (2013). One complicated extended family: the influence of alphabetic knowledge and vocabulary on phonemic awareness. *J. Res. Read.* 36, 29–41. doi: 10.1111/j.1467-9817.2010.01486.x

Perfetti, C. A., Beck, I., Bell, L. C., and Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: a longitudinal study of first grade children. *Merrill Palmer Q.* 33, 283–319. Available online at: www.jstor.org/stable/23086537.

Preston, J. L., Molfese, P. J., Frost, S. J., Mencl, W. E., Fulbright, R. K., Hoeft, F., et al. (2016). Print-speech convergence predicts future reading outcomes in early readers. *Psychol. Sci.* 27, 75–84. doi: 10.1177/0956797615611921

Price, C. J., and Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253. doi: 10.1016/j.tics.2011.04.001

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., et al. (2001). Neurobiological studies of reading and reading disability. *J. Commun. Disord.* 34, 479–492. doi: 10.1016/S0021-9924(01)00060-0

Richlan, F. (2019). The functional neuroanatomy of letter-speech sound integration and its relation to brain abnormalities in developmental dyslexia. *Front. Hum. Neurosci.* 13:21. doi: 10.3389/fnhum.2019.00021

Richlan, F., Kronbichler, M., and Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 30, 3299–3308. doi: 10.1002/hbm.20752

Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568. doi: 10.1037/0033-295x.96.4.523

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition* 55, 151–218. doi: 10.1016/0010-0277(94)00645-2

Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biol. Psychiatry* 57, 1301–1309. doi: 10.1016/j.biopsych.2005.01.043

Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., et al. (2003). Neural systems for compensation and persistence: young adult outcome of childhood reading disability. *Biol. Psychiatry* 54, 25–33. doi: 10.1016/s0006-3223(02)01836-x

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., et al. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol. Psychiatry* 52, 101–110. doi: 10.1016/s0006-3223(02)01365-3

Simos, P. G., Breier, J. I., Wheless, J. W., Maggio, W. W., Fletcher, J. M., Castillo, E. M., et al. (2000). Brain mechanisms for reading: the role of the superior temporal gyrus in word and pseudoword naming. *Neuroreport* 11, 2443–2446. doi: 10.1097/00001756-200008030-00021

Stevenson, R. A., and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034

Stuart, M. (1990). Processing strategies in a phoneme deletion task. *Q. J. Exp. Psychol.* 42, 305–327. doi: 10.1080/14640749008401224

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang.* 9, 182–198. doi: 10.1016/0093-934x(80)90139-x

Temple, E., Poldrack, R. A., Salidis, J., Deutsch, G. K., Tallal, P., Merzenich, M. M., et al. (2001). Disrupted neural responses to phonological and orthographic processing in dyslexic children: an fMRI study. *Neuroreport* 12, 299–307. doi: 10.1097/00001756-200102120-00024

Torgesen, J. K., Rashotte, C. A., and Wagner, R. K. (1999). *TOWRE: Test of Word Reading Efficiency.* Austin, TX: Pro-Ed, Inc.

Tsapkini, K., and Rapp, B. (2010). The orthography-specific functions of the left fusiform gyrus: evidence of modality and category specificity. *Cortex* 46, 185–205. doi: 10.1016/j.cortex.2009.02.025

Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., and Eden, G. F. (2003). Development of neural mechanisms for reading. *Nat. Neurosci.* 6, 767–773. doi: 10.1038/nn1065

Vaessen, A., Gerretsen, P., and Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: double deficits explored. *J. Exp. Child Psychol.* 103, 202–221. doi: 10.1016/j.jecp.2008.12.004

van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025

van Atteveldt, N. M., Roebroeck, A., and Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear. Res.* 258, 152–164. doi: 10.1016/j.heares.2009.05.007

Wagner, R. K., and Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychol. Bull.* 101, 192–212. doi: 10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., and Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing: CTOPP.* Austin, TX: Pro-Ed.

Wechsler, D. (2011). *WASI-II: Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson Tests of Achievement.* Itasca, IL: Riverside Publishing.

Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., and Bonte, M. (2015). Crossmodal deficit in dyslexic children: practice affects the neural timing of letter-speech sound integration. *Front. Hum. Neurosci.* 9:369. doi: 10.3389/fnhum.2015.00369

Zeno, S., Ivens, S., Millard, R., and Duvvuri, R. (1995). *The Educator's Word Frequency Guide (S. Zeno Ed.).* Brewster, NY: Touchstone Applied Science Associates.

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

SYSTEMATIC REVIEW
published: 12 March 2020
doi: 10.3389/fnhum.2020.00062

Functional Brain Connectivity of Language Functions in Children Revealed by EEG and MEG: A Systematic Review

Isabelle Gaudet[1,2]*, *Alejandra Hüsser*[1,2], *Phetsamone Vannasing*[1] *and Anne Gallagher*[1,2]*

[1] *Laboratoire d'imagerie optique en neurodéveloppement (LIONLAB), Sainte-Justine University Hospital Research Center, Montréal, QC, Canada,* [2] *Department of Psychology, Université de Montréal, Montréal, QC, Canada*

Edited by:
Claude Alain,
Rotman Research Institute (RRI), Canada

Reviewed by:
Jascha Ruesseler,
University of Bamberg, Germany
Sam Doesburg,
Simon Fraser University, Canada

***Correspondence:**
Isabelle Gaudet
isabelle.gaudet.3@umontreal.ca
Anne Gallagher
anne.gallagher@umontreal.ca

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *03 October 2019*
Accepted: *10 February 2020*
Published: *12 March 2020*

Citation:
Gaudet I, Hüsser A, Vannasing P and Gallagher A (2020) Functional Brain Connectivity of Language Functions in Children Revealed by EEG and MEG: A Systematic Review. Front. Hum. Neurosci. 14:62. doi: 10.3389/fnhum.2020.00062

The development of language functions is of great interest to neuroscientists, as these functions are among the fundamental capacities of human cognition. For many years, researchers aimed at identifying cerebral correlates of language abilities. More recently, the development of new data analysis tools has generated a shift toward the investigation of complex cerebral networks. In 2015, Weiss-Croft and Baldeweg published a very interesting systematic review on the development of functional language networks, explored through the use of functional magnetic resonance imaging (fMRI). Compared to fMRI and because of their excellent temporal resolution, magnetoencephalography (MEG) and electroencephalography (EEG) provide different and important information on brain activity. Both therefore constitute crucial neuroimaging techniques for the investigation of the maturation of functional language brain networks. The main objective of this systematic review is to provide a state of knowledge on the investigation of language-related cerebral networks in children, through the use of EEG and MEG, as well as a detailed portrait of relevant MEG and EEG data analysis methods used in that specific research context. To do so, we have summarized the results and systematically compared the methodological approach of 24 peer-reviewed EEG or MEG scientific studies that included healthy children and children with or at high risk of language disabilities, from birth up to 18 years of age. All included studies employed functional and effective connectivity measures, such as coherence, phase locking value, and Phase Slope Index, and did so using different experimental paradigms (e.g., at rest or during language-related tasks). This review will provide more insight into the use of EEG and MEG for the study of language networks in children, contribute to the current state of knowledge on the developmental path of functional connectivity in language networks during childhood and adolescence, and finally allow future studies to choose the most appropriate type of connectivity analysis.

Keywords: functional connectivity, cerebral networks, language, language development, children, EEG, MEG, connectivity analysis

INTRODUCTION

Language is a highly complex function that is importantly involved in the development of human cognition and social functions (Berwick et al., 2013). With major advances in neuroimaging techniques, the language neural architecture has been increasingly studied in the past 20 years. While several brain regions have been identified as key areas for expressive and receptive language, it is now also widely recognized that the latter relies more on complex neural networks, requiring coordination between distinct neuronal populations and less on independent and specific brain areas (Ardila et al., 2015; Tremblay and Dick, 2016).

Over the past decades, functional brain connectivity (FC) has progressively captured the interest of scientists and clinical researchers working in the field of cognitive neuroscience, leading to the publication of numerous articles on the subject. On a general note, functional connectivity is defined as the statistical relationships between cerebral signals over time and thus potentially allows conclusions to be made regarding the functional interactions between two or more brain regions. Effective connectivity, on the other hand, goes beyond the correlations between cerebral activity and aims at specifying causal relationships through the use of experimental paradigms or models. This allows for an interpretation of the direction of interactions between different cerebral regions (Friston, 2011). With the sharp increase of studies on brain connectivity, researchers have developed and applied increasingly sophisticated analytic strategies that highlight functional or effective connectivity (EC) and that allow a more advanced exploration of interactions between regional structures and networks involved in language development (Bastos and Schoffelen, 2016). In the past few years, novel neuroimaging techniques and methods of analysis have enabled the examination of functional connectivity patterns. Namely, functional magnetic resonance imaging (fMRI) was the neuroimaging technique used in the first published study of brain spontaneous fluctuations, measured at rest (Biswal et al., 1995). Functional magnetic resonance imaging is widely used in brain connectivity studies, mostly due to its high spatial resolution (in millimeters). However, because it relies on the coupling between cerebral blood flow (hemodynamic response) and the underlying neuronal activation, this technique provides only an indirect measure of brain activity. Moreover, even though neuronal events occur within milliseconds, the induced blood-oxygenation changes spread out over several seconds, thereby severely limiting fMRI's temporal resolution (∼2–3 s). Techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), on the other hand, provide direct information on neuronal electrical activity and offer higher temporal resolution (<1 millisecond). This is particularly relevant for the study of language functions, because auditory processing and language processing occur within a short time interval of milliseconds (Skeide and Friederici, 2016).

So far, neuronal accounts of language system development largely rely on EEG data (Skeide and Friederici, 2016). Traditionally, electrophysiological data have been examined for event-related potential (ERP), a method that reflects the brain's activity in response to a particular stimulus event. As of now, several metrics can be used to estimate functional connectivity between electrodes.

In order to perform functional connectivity analysis, MEG and EEG (M/EEG) data are commonly transformed into the frequency domain. Measures are thus typically classified by five fundamental frequency bands, mostly defined by their spectral boundaries: delta (<4 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz), and gamma (>30 Hz) (Cacioppo et al., 2007), each of which has different functional characteristics and cortical topography (Herrmann et al., 2016). Despite the fact that the definitions of these bands may vary between studies, and the boundaries used in studies of early childhood may be lower (Saby and Marshall, 2012), the interpretation arising from the present systematic review is based on the above definition by Cacioppo et al. (2007).

What is more, development and maturation affect the frequency and synchronization of neural oscillations, both at rest and during a cognitive task. Globally, analyses of resting state networks reveal that slow-wave activity (delta and theta) tends to decrease throughout childhood and adolescence, whereas oscillations in higher frequency (alpha, beta, and gamma) show an increase with age (Uhlhaas et al., 2010). Moreover, FC in childhood is dominated by short-distance local links, which are gradually replaced by long-distance functional connections in adulthood, thus forming mature cerebral networks (Vértes and Bullmore, 2015; Meng and Xiang, 2016; Oldham and Fornito, 2018). The task-related developmental trajectory of neural oscillations is, however, less clear and varies widely depending on the nature of the task.

When it comes to the functional meaning of different frequency bands, previous studies have suggested that brain signals of each frequency band play a different role. First, the coherence of local neuronal populations and bottom-up processing are associated with high-frequency oscillations (Buzsáki et al., 2013; Friederici and Singer, 2015). Slower frequency ranges, on the other hand, are understood to represent the cooperative activity of large-scale neuronal networks and mediate top-down feedback information (Palva and Palva, 2018).

Regarding language processing, the use of FC in the spectral domain is certainly important, but little is known about the association between frequency bands and language networks. Nevertheless, distinctions have been made regarding language processing and frequency band using spectral power analyses. It is argued that different stages of auditory and speech processing, language comprehension, and active speech itself do not rely on the same frequency bands (for an exhaustive review see Kösem and Van Wassenhove, 2017; Meyer, 2018). More specifically, delta range (<4 Hz) has been associated with intonational processing and syntactic comprehension (Kösem and Van Wassenhove, 2017; Meyer, 2018). It plays a role in top-down processing and seems to contribute to the organization of the cortical speech system, which regulates auditory-cortical excitability. It is further implicated in language comprehension, more precisely in the grouping of words into syntactic phrases (Meyer, 2018). It has been pointed out that theta (4–7 Hz) synchronizes with syllabic rates (Giraud and Poeppel, 2012; Meyer, 2018) and that theta coherence increases in tasks involving verbal information retrieval and verbal working memory (Friederici and Singer, 2015; Meyer, 2018). Alpha (8–12 Hz) oscillations may also play a role in verbal working memory (Friederici and Singer, 2015; Meyer, 2018). Beta activity (13–30 Hz) in language processing has been associated with semantic predictions (top-down mechanisms), as well as in syntactic and semantic binding mechanisms. It has also been correlated with verbal memory processes and language production (Weiss and Mueller, 2012). Finally, the gamma band (>30 Hz) has been associated with phonological perception and assessment of the contextual semantic fit of incoming words [bottom-up; (Meyer, 2018)]. The association of functional connectivity based on frequency bands and the different stages of language processing are still subject to investigation.

Several techniques have been proposed in order to measure cerebral activity, thus allowing for the interpretation of brain connectivity. Even though a large range of FC metrics is available in the current literature, the present article is limited to those brain connectivity approaches used in pediatric electrophysiological language research. Thus, FC analysis will not be addressed exhaustively. Only the most commonly used metrics to quantify brain connectivity, such as coherence, phase locking value (PLV), Phase Lag Index (PLI), correlation, Granger causality, and Graph theory, will be briefly described in this review. Complementary reviews on more detailed mathematical analyses of connectivity methods can be consulted elsewhere (e.g., Kida et al., 2015; Bastos and Schoffelen, 2016).

Connectivity analyses in M/EEG traditionally include the examination for changes in coherence between sources or sensors. Coherence can be defined as the covariation in amplitude and phase between two signals and quantifies the linear correlation between two time series, and this on the frequency domain (Bowyer, 2016). It is assumed that the higher the correlation, the more synchronized, and therefore integrated, the signals are. Thus, coherence is sensitive to changes in both power and phase relationships but cannot provide direct information on the true relationship between the two signals (Sakkalis, 2011).

As an alternative to traditional amplitude-based indices of coherence, metrics of phase synchronization have been developed, such as PLV and PLI. Both PLV and PLI compute the consistency of phase difference between two variables over a time period. They provide a measure of the two signals' temporal relationship, independent of their signal amplitude (Lachaux et al., 1999). The PLV approach evaluates the instantaneous phase difference of signals, assuming that the connected areas generate signals whose phases evolve together. Therefore, the phases of the signals are considered synchronous or locked if the difference between them is constant (Bruña et al., 2018). Similarly, PLI estimates the asymmetry of the distribution of phase differences between two signals, but this method is designed to reduce the effect of volume conduction (Stam et al., 2007). The central idea is that a consistent phase difference between two times series (asymmetric distribution, $\mathrm{PLI} > 0$), cannot result from a single source (volume conduction). Overall, phase synchronization metrics are better used for short-duration events such as in event-related studies, to determine the coupling of two signals across trials (Aydore et al., 2013; Bowyer, 2016).

Recently, directed connectivity or EC metrics have been developed to determine the nature of the neural interactions that enable information flux, such as Granger causality in the time domain (Bressler and Seth, 2011) or phase slope index (PSI) in frequency domain (Nolte et al., 2008). Based on phase differences, PSI is a weighted average measure of phase coherency slope between two signals, over a frequency band (Nolte et al., 2008; Bastos and Schoffelen, 2016). Some EC measures rely on the concept of Granger causality, whereby one time series is said to "Granger cause" a second one if the past values of the first improve the prediction of the second. Originally, the concept of Granger causality was applied to time series, but this approach has been extended to the frequency domain (Geweke, 1982), and many multivariate measures can be derived from this model (Sakkalis, 2011).

Similar to fMRI or other neuroimaging techniques, M/EEG data used along with connectivity matrices can be used to construct brain networks from FC measures of the frequency domain (PLI, PLV, coherence), the source space domain, or the EC models (Sporns et al., 2004; Stam, 2004; Bullmore and Sporns, 2009). Subsequent connectivity metrics of all paired electrodes can then be explored between regions, using the Graph theory approach (Stam and Van Straaten, 2012). This method represents the brain as a collection of nodes, corresponding to recording sites or brain regions, and the pairwise relationship between them (edges). Taken together, nodes and edges enable the quantitative description of the local and global topological organization of brain networks (Van Diessen et al., 2015). It has been shown that small-world topology is found at different frequency bands (Stam, 2004) and can be associated with cognitive performance and developmental changes in functional brain networks in young children (Boersma et al., 2013).

Despite the growing number of published studies on language brain connectivity, the establishment of functional patterns of language networks during childhood and adolescence is not yet fully understood. In 2015, Weiss-Croft and Baldeweg (2015) published the first and only systematic review of studies that used fMRI to explore the development of functional language networks. The authors identified both progressive (increasing) changes of FC with age, associated with cerebral specialization, and regressive (decreasing) changes of FC with age, associated with more automatized language processing and lower engagement of control mechanisms (Weiss-Croft and Baldeweg, 2015). Specifically, their review highlights four main findings. First, brain activity in regions that support semantic processing increased throughout development, reflecting specialization of the brain. Second, with age, there is an increased activation in sensory–motor regions, along with a decreased activation in higher–order cognitive regions. Third, an age-related decreased activation was found in regions implicated in the default mode network (posterior cingulate cortex and precuneus). Finally, their results demonstrate the establishment of language lateralization by the age of 5 years. Although this study is indeed interesting, there is currently in the literature no systematic review that includes M/EEG studies. Because of the excellent temporal resolution of MEG and EEG, such a study would greatly help to provide additional and important information on the establishment of functional patterns of language networks. Therefore, the main objectives of this article are to provide a state of knowledge on the investigation of language-related cerebral networks in children, through the use of M/EEG, and a detailed portrait of relevant M/EEG data analyses methods that have been used in the assessment of language functional connectivity in children. To do so, we conducted this systematic review on functional, and to some extent effective, connectivity patterns of spoken language in children, as revealed by EEG or MEG. Given the multitude of metrics used to quantify oscillatory interactions (e.g., coherence, phase locking, connectivity matrices, graph theory, PSI) and the diversity of methodological designs (e.g., resting state vs. task recording, large variety of language tasks, longitudinal vs. cross-sectional study), the secondary objective is to synthesize and compare various method of connectivity analysis in the context of different pediatric populations (healthy and clinical) and a wide range of research objectives.

METHODS

Search Strategy

The literature review was conducted using five databases: PubMed, PsycINFO, Web of Science, Scopus, and Linguistics and Language Behavior in order to find articles published between January 1995 and June 2018 inclusively. The key terms used were as follows: (magnetoencephalography OR electroencephalography OR MEG OR EEG) AND (resting state OR functional connectivity OR synchron∗ OR network∗ OR effective connectivity OR coherence) AND (Language OR Speech) AND (infant∗ OR infancy OR child OR children OR youth∗ OR toddler∗ OR schoolchild∗ OR teenager∗ OR adolescent∗ OR kid OR kids OR newborn). Additional reports were identified by handsearching the references cited in the retrieved articles.

Selection Criteria

This review is limited to empirical studies published in peer-reviewed journals in English or in French. Studies that adhered to the following inclusion criteria were considered: (1) The study included children or adolescents (<18 years old), although the age range may extend into adulthood; (2) functional or EC analysis was performed based on EEG or MEG data. We verified whether the described methods allowed actual interpretation of functional connectivity or applied different techniques such as intertrial synchronization, ERP timing, or time-frequency analysis, which were sometimes referred to as functional connectivity, but do not in fact fall in this category (Sakkalis, 2011; Bastos and Schoffelen, 2016). (3) Studies that investigated language networks were included if either one of the following two conditions was met: (a) the authors used a behavioral assessment before or after the imaging acquisition, in order to evaluate language abilities; or (b) the authors applied expressive or receptive language paradigms (e.g., speech stimuli, story listening, or speech production) during MEG or EEG recording. In order to provide an exhaustive view of the connectivity patterns associated with language in childhood, this systematic review includes clinical pediatric samples as well as healthy children, as long as the methodology fit our selection criteria. Articles about written language only (reading or writing) without any association with verbal comprehension or expressions have been excluded.

The lists of references of the selected articles were searched manually for additional relevant articles. The study selection process is summarized in **Figure 1**.

Data Extraction

Following the database search, duplicates were removed. For all remaining articles, titles and abstracts were reviewed by the first author (IG) and selected for a second revision if they met at least one of the inclusion criteria. For the second revision, remaining articles were reviewed independently by two authors (IG and AH), in order to determine whether they matched the purpose of this study. When no consensus was reached, the consultation of a third-party expert in the domain (PV) helped make the ultimate decision on eligibility. **Figure 1** shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) workflow diagram for study selection. Relevant information from each article was entered into a spreadsheet that included: (1) sample characteristics: age, gender, IQ, language evaluation method, sample size; (2) experimental paradigms: resting state, event-related experiments, sleep studies; (3) brain recording technique (EEG or MEG); (4) connectivity metrics.

The wide variability in study characteristics along these methodological dimensions precluded a meta-analysis. Instead, we synthesized and critically appraised findings made through the use of functional connectivity in the study of spoken language in children.

RESULTS

A total of 704 articles were screened in the first step. Of these, 507 were excluded on the basis of their title or abstract, either because they were not experimental studies (e.g., review), they were conducted with adult participants only, or they did not conduct connectivity analysis using EEG or MEG. Following these exclusions, 197 articles were assessed for eligibility. Of these, 173 were excluded because they did not meet at least one of the selection criteria.

A total of 24 articles met the selection criteria, passed interrater revision (79% agreement), and were confirmed by the third-party expert. All publications included in this work are peer-reviewed studies about FC of language functions in children, as revealed by EEG or MEG, and were published between 1999 and 2018. Detailed information was gathered about each study's population of interest, sample size, age of participants, design, imaging paradigm, type of language assessment, frequency bands considered for analyses, use of source or sensor analyses, and, finally, approach for connectivity analysis (see **Table 1** for studies including healthy children and **Table 2** for those addressing clinical populations). Each table begins with studies

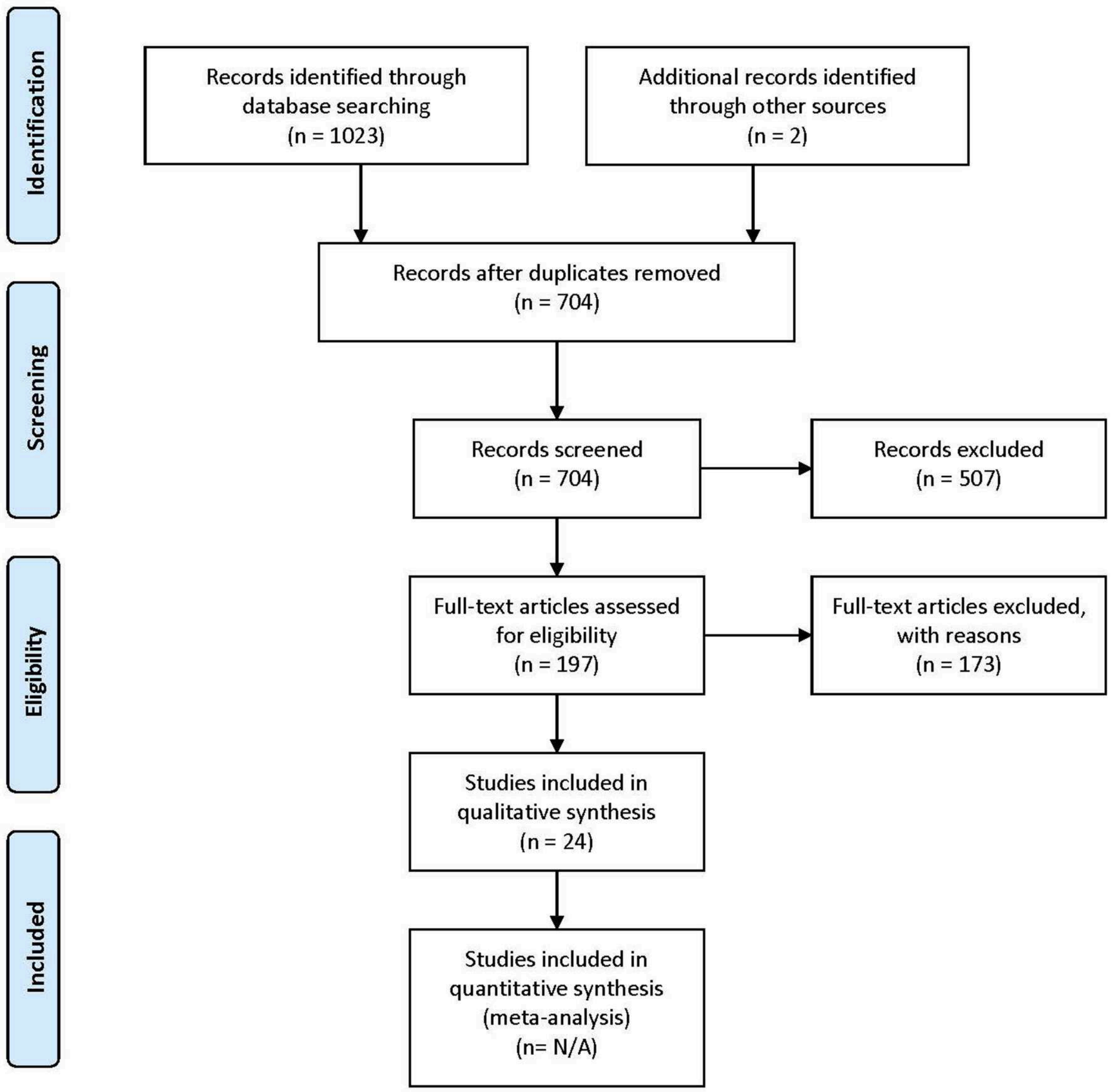


FIGURE 1 | PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram describing the paper selection process. Figure adapted from Moher et al. (2009).

using EEG (**Tables 1A**, **2A**) followed by those employing MEG (**Tables 1B**, **2B**).

Thirteen of the articles covered in this review addressed functional connectivity and language functions in healthy children, whereas 11 included children at risk of or suffering from various clinical conditions. **Table 3** shows the different populations included in these studies. The most studied pathologies were related to language impairments such as dyslexia, language learning disorders, and stuttering (20%), as well as autism spectrum disorder (ASD) (13%).

Figure 2 shows the distribution of the number of participants per age group taken together for all studies, both in healthy and clinical populations. Infancy includes the first year after birth (0–12 months). Toddlers are children aged between 1 and 3 years; preschoolers include the period from 3 to 5 years of age, grade-schoolers from 5 to 12 years, and adolescents are participants between 12 and 18 years of age. Each age group is subdivided into the number of children included in the studies addressing various clinical populations (green bars) and those interested in healthy children (blue bars), including those used as controls. Most of the healthy children studied were toddlers ($n > 350$), whereas studies interested in the impact of pathological conditions mostly included grade-schoolers ($n > 150$), although several studies on clinical populations also included infants and preschoolers. No data were available for any toddler or adolescent populations with clinical conditions. Overall, studies included in this systematic review total together a sample size of 728 in studies of healthy children and 394 in studies of clinical populations.

TABLE 1A | Descriptive data and methodological outline of articles focusing on healthy children in EEG studies.

References	*n* (M/F)	Age	Design	EEG/MEG paradigm	Language assessment	Frequency band(s)	Source/ sensor	Connectivity analysis
EEG								
Asano et al., 2015	13/6	11 mo	Cross-sectional	Symbol–sound mismatch	N/A	Alpha, beta	Sensor	Phase locking value
Hanlon et al., 1999	284/224	0–16.75 y	Cross-sectional	Resting	N/A	Theta	Sensor	Coherence
Kühn-Popp et al., 2016	15/17	14; 15 and 42 mo	Longitudinal	Resting	Declarative pointing and Verbal-IQ	Theta–alpha	Sensor	Coherence
Marshall et al., 2008	48/42	30 and 42 mo	Longitudinal	Resting	Reynell Developmental Language Scales	Theta, alpha, beta	Sensor	Coherence
Mundy et al., 2003	18/14	14–24 mo	Longitudinal	Resting	MCDI	Theta	Sensor	Coherence
Poblano et al., 2016	18/18	9–16 y	Cross-sectional	Resting; Lexical-tonal discrimination	N/A	Theta	Sensor	Pearson correlation
Whedon et al., 2016	153/147	6–34 mo	Longitudinal	Resting	PPVT-III^2	Theta–alpha	Sensor	Coherence
Yang et al., 2005	23 (N/A)	6–8 y	Cross-sectional	Resting	Verbal-IQ	Delta, theta, alpha, beta	Sensor	Pearson correlation

TABLE 1B | Descriptive data and methodological outline of articles focusing on healthy children in MEG studies.

References	*n* (M/F)	Age	Design	EEG/MEG paradigm	Language assessment	Frequency band(s)	Source/ sensor	Connectivity analysis
MEG								
Doesburg et al., 2016	31/42	4–18 y	Cross-sectional	Word generation	PPVT, EVT	Alpha, beta, theta	Source	Phase locking value, phase lag index, graph theory
Doesburg et al., 2012	5/5	16–19 y	Cross-sectional	Word generation	N/A	Gamma, theta	Source	Phase locking value
Kadis et al., 2016	13/8	5–18 y	Retrospective	Word generation	N/A	All	Source	Phase slope index
Kikuchi et al., 2011	36/42	32–64 mo	Cross-sectional	Story listening	Expressive Vocabulary and Riddles (K-ABC)	Delta, theta, alpha, beta	Sensor	Coherence
Youssofzadeh et al., 2017	13/16	4–18 y	Cross-sectional	Word generation	N/A	Theta, alpha, beta, gamma	Source	Phase locking value

Studies in the first part of the table used EEG, whereas those in the second part applied MEG.
M, male; F, female; N/A, not applicable; MCDI, Mac-Arthur communicative developmental inventory; PPVT, Peabody Picture Vocabulary Test; ECT, Expressive Vocabulary Test; K-ABC, Kaufman Assessment Battery.

Different methods of connectivity analyses were used in these studies; they are summarized in **Table 4**. Some studies combined or compared several methods for estimating cerebral connectivity. Phase coherence analysis was the most common method used (45%), followed by PLV (21%). The analyses were based on all frequency bands, as specified in **Tables 1**, **2**. The most studied frequency band was theta, and the least studied was gamma. Sixteen studies used sensor information, and seven applied a source analysis. One study reported results for both source and sensor-based analyses.

Despite the fact that the scope of these studies differed, the aim of this review is to capture common findings concerning language-related functional connectivity. Therefore, we first present an overview of the results that emerged from the studies that investigated the association between language functioning and connectivity patterns, regardless of the task used during the EEG or MEG recording. Second, we illustrate, separately for healthy children and those in clinical populations FC and EC findings, while an expressive or receptive task was performed during the EEG and MEG recording. Finally, we display the results that emerge from all included studies organized according to the types of connectivity analyses used, beginning with those using functional connectivity, followed by those using EC. Again, the results will be indicated separately for healthy children and children with various clinical conditions.

Overview of All Results

From the 24 articles included in the review, only nine attempted to associate FC or EC patterns with objective measures of language functioning. **Figure 3** shows the main results from these

TABLE 2A | Descriptive data and methodological outline of articles focusing on children with or at risk of different clinical conditions in EEG studies.

References	Pathology	*n* (M/F)	Age	Design	EEG/MEG paradigm	Language assessment	Frequency band(s)	Source/ sensor	Connectivity analysis
EEG									
Righi et al., 2014	Risk of autism	54 (N/A)	6 and 12 mo	Longitudinal	Discrimination of consonants	Subtest of Mullen Scales of Early Learning	Gamma	Sensor	Coherence
Njiokiktjien et al., 2001	Nonverbal learning disorder/ Language disorder[1]	12/6 12/6	6–11 y	Cross-sectional	Resting	N/A	All	Sensor	Coherence
Zare et al., 2016	Risk of language disorder[1]	17/7	6 mo	Cross-sectional	Resting	N/A	Delta, theta, alpha1, alpha2	Sensor	Connectivity matrix, graph theory
Kabdebon et al., 2015	Prematurity/ healthy	18/12 10/5	8 mo	Cross-sectional	Syllabic learning	N/A	Alpha, beta	Sensor	Coherence
Vasil'yeva and Shmalei, 2013	Stammering/ healthy	47/0 59/0	3–5 y	Cross-sectional	Resting	N/A	All	Sensor	Coherence
Williams et al., 2012	Congenital heart disease	14/2	0–18 mo	Longitudinal	Resting	Bayley Scales of Infant Development	Beta	Sensor	Coherence

TABLE 2B | Descriptive data and methodological outline of articles focusing on children with or at risk of different clinical conditions in MEG studies.

References	Pathology	*n* (M/F)	Age	Design	EEG/MEG paradigm	Language assessment	Frequency band(s)	Source/ sensor	Connectivity analysis
MEG									
Kovelman et al., 2015	Autism/healthy	10 (N/A) 9 (N/A)	8–12 y	Cross-sectional	Discrimination of native and foreign language	N/A	All	Source	Coherence
Mamashli et al., 2017	Autism/healthy	29/0 17/0	9–15 y	Cross-sectional	Tonal discrimination	Social communication questionnaire	All	Source	Coherence
Molinaro et al., 2016	Dyslexia/healthy	9/11 10/10	8–14 y	Cross-sectional	Sentence listening	Verbal fluency, rapid automatized naming, pseudoword repetition, and phonemic deletion	Delta, theta	Sensor, Source	Coherence, partial direct coherence based on Granger causality
Lizarazu et al., 2015	Language disorder[a]/healthy	6/4 5/5	8–14 y	Cross-sectional	Listening of sounds	Reading of word and pseudoword lists, pseudoword repetition, and phonemic deletion	Delta, theta, beta, and gamma	Source	Phase locking value
Barnes-Davis et al., 2018	Extreme prematurity/term born	9/6 7/8	4–6 y	Cross-sectional	Story listening	PPVT, Expressive Vocabulary Test	Beta	Sensor	Phase slope and phase lag index

Studies in the first part of the table used EEG, whereas those in the second part applied MEG.
[a] Language-based learning disorders (e.g., dyslexia, dysphasia).
M, male; F, female; N/A, not applicable; PPVT, Peabody Picture Vocabulary Test.

nine studies, for healthy subjects (eight studies) and for a clinical population (one study). Results are presented for each frequency band and organized according to age.

Only one study (Williams et al., 2012) investigated the relationship between FC networks and language abilities in a clinical population, that is, children with congenital heart disease (CHD), who are known to be at high risk of language delay (Hövels-Gürich et al., 2008; Hövels-Gürich and Mccusker, 2016; Fourdain et al., 2019). The authors did not find any significant association between FC during the neonatal period and their later language abilities as measured at 18 months of age. Additionally, Marshall et al. (2008) found no significant correlation between FC patterns and language performance in preschoolers under foster care. However, seven studies found a significant relationship between FC in the theta band and language performance. Positive correlations between FC and language score were also found in higher frequency bands: alpha (Yang et al., 2005; Doesburg et al., 2016) and beta (Yang et al., 2005; Doesburg et al., 2016). It should be noted that no study investigated the relationship between language skills and FC patterns in the gamma band.

In addition to articles that included a behavioral assessment of language functions, performed before or after an EEG or MEG recording, this systematic review also considers studies that included an expressive or receptive language paradigm (e.g., speech stimuli or speech production) during an MEG or EEG recording. The FC or EC patterns that arose from language paradigms are summarized in **Figure 4** (for healthy children) and **Figure 5** (for clinical populations).

In healthy children, the use of an expressive language paradigm (usually a verb generation task) was favored in four studies, whereas three studies used a receptive language task in order to examine the connectivity patterns that underlie language processing. These types of research paradigms have been performed mostly in research pertaining to grade-schoolers and adolescents, and the results are spread across all frequency bands.

In clinical populations, language tasks were mainly used to compare FC patterns between vulnerable children and healthy children. Here, only receptive language paradigms were used during M/EEG recording. Differences in FC between healthy and clinical subjects occur predominantly in the higher frequency bands (beta and gamma). Again, more details on the results of these studies are provided in section Results Derived From Connectivity Metrics.

Finally, it should be noted that two studies (Njiokiktjien et al., 2001; Vasil'yeva and Shmalei, 2013) done in resting-state FC in clinical populations were not presented in any of these figures. One of these studies looked at FC in children who received a diagnosis of language-based learning disorder (LLD), compared to children with non-verbal learning disorders

TABLE 3 | Overall composition of samples included in all studies.

Study population	% (*n*)
Healthy	54 (13)
Autism spectrum disorder	13 (3)
Prematurity	9 (2)
Dyslexia	8 (2)
Language learning disorders	8 (2)
Stuttering	4 (1)
Congenital heart disease	4 (1)

TABLE 4 | Overview of all approaches applied to analyze functional or effective connectivity in included studies.

Connectivity analysis	% (*n*)*
Coherence	45 (13)
Phase locking value	21 (6)
Pearson correlation	7 (2)
Graph theory	7 (2)
Phase slope index	7 (2)
Phase lag index	7 (2)
Connectivity matrices	3 (1)
Granger causality	3 (1)

**Some studies applied multiple analyses; hence the total n outranges the number of studies included in this review.*

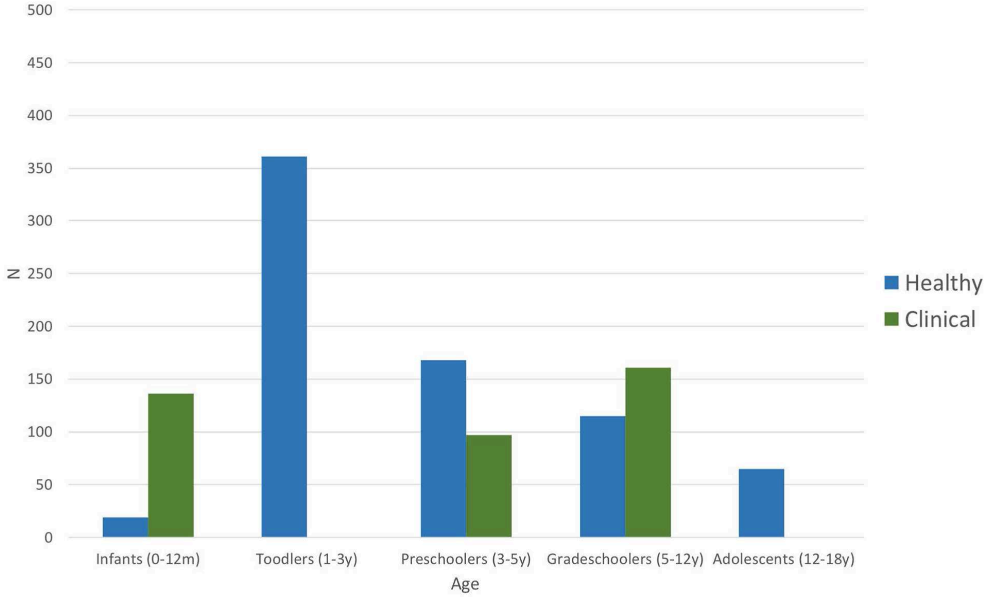


FIGURE 2 | Number of participants per age group of all included studies ($n = 24$). Blue bars represent number of participants included in the articles addressing healthy children; green bars stand for the number of participants included in studies investigating clinical populations (including control groups) such as autism spectrum disorder, dyslexia, language-learning impairment, or prematurity (**Table 3**).

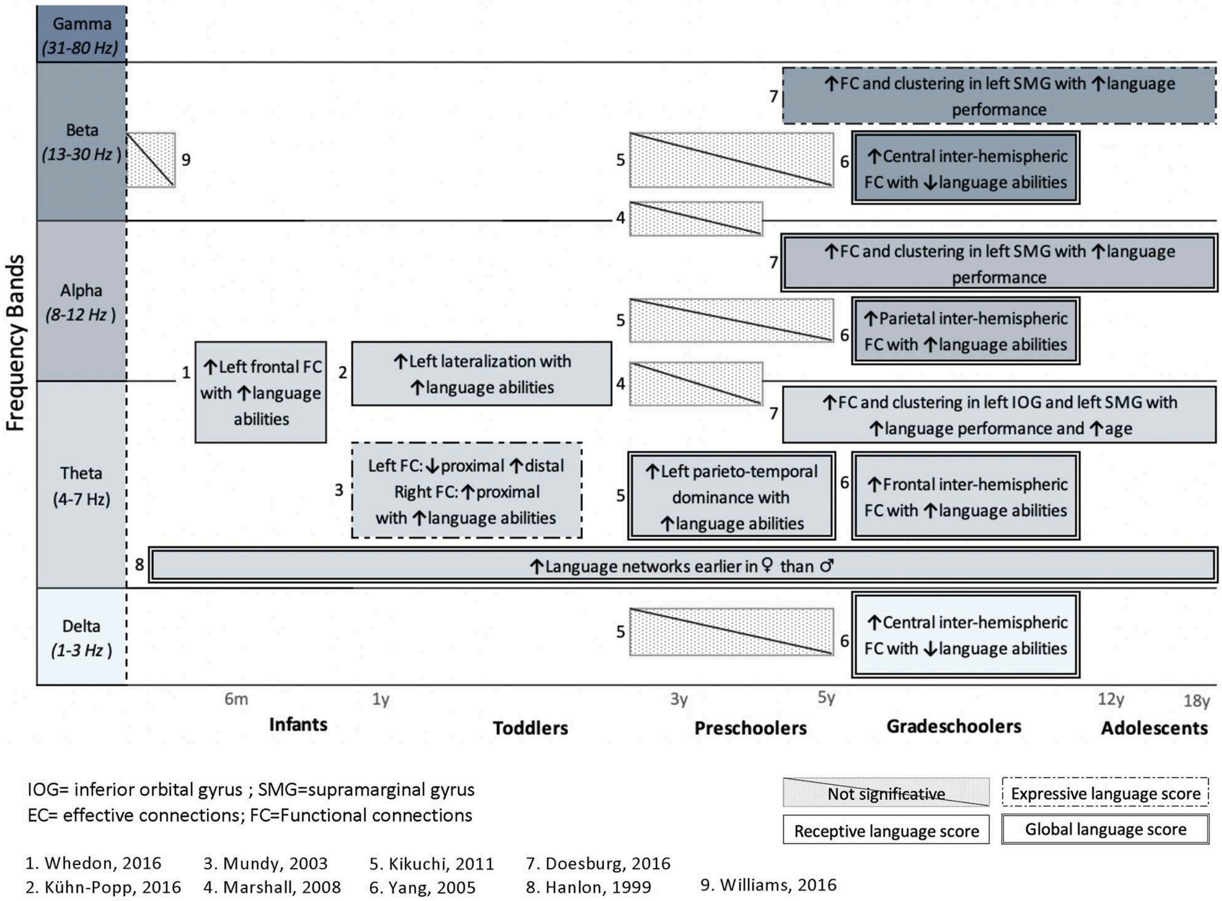


FIGURE 3 | Summary of studies investigating the association between language abilities, assessed with standardized tools, and cerebral language networks. Results are presented for each frequency band and organized regarding ages. Studies in healthy subjects ($n = 8$) and a clinical population ($n = 1$) are included. Upper arrows (↑) indicate a positive correlation with either receptive (simple solid line), expressive (dashed lines 1), or global language functioning (solid double lines), whereas downward arrow (↓) indicates negative correlation with language. Hatched areas represent non-significant correlations with language abilities.

(Njiokiktjien et al., 2001). The other looked at the FC patterns in children who stutter (Vasil'yeva and Shmalei, 2013). These studies did not use a language paradigm during EEG recording and therefore do not directly correlate connectivity patterns with behavioral language measures. The results of these two studies will nonetheless be discussed in section Results From Coherence in Clinical Population.

Results Derived From Connectivity Metrics

Results From Correlation and Coherence Analyses

The correlation coefficient and its analog in the frequency domain, coherence, are the classic measures of interdependence between two signals (Sakkalis, 2011; Van Mierlo et al., 2014; Hassan and Wendling, 2018). Based on the amplitudes of the signals, the cross-correlation coefficient is a measure of the linear correlation between two time series and was utilized in one study using a tonal discrimination task (Poblano et al., 2016). Coherence, on the other hand, detects the linear relation between two electrophysiological signals at any particular frequency (Van Mierlo et al., 2014; Bowyer, 2016). It is mainly used at rest and appears to be the most popular metric for M/EEG evaluation of functional language networks in children ($n = 13$). One other study used coherence and Granger causality and will therefore be discussed in the section on EC.

Results from correlation in healthy children

In a study on adolescents (9–16 years old, Poblano et al., 2016), correlation analyses were performed between several recording sites of the brain and were acquired during a lexical-tonal discrimination task of bisyllabic words in the Zapotec language (a tonal language, spoken by the participants). Results showed significant increases of interhemispheric and intrahemispheric correlations of the theta-relative power during a word discrimination task, predominantly between left frontal and right temporal sites.

Results from coherence in healthy children

In healthy infants, few studies ($n = 6$) investigated the association between measures of coherence and later language abilities of preschoolers (Mundy et al., 2003; Marshall et al., 2008; Kikuchi

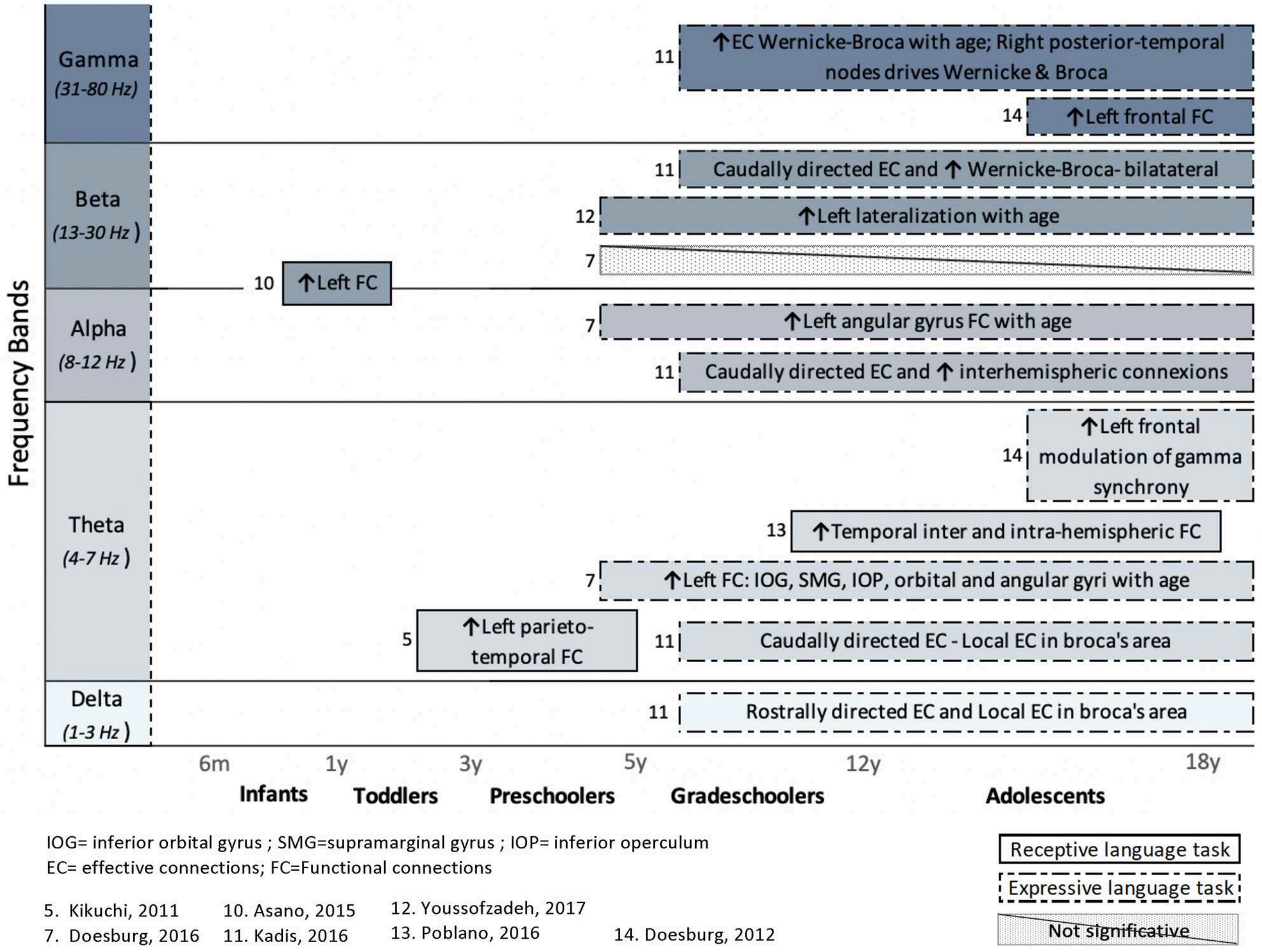


FIGURE 4 | Overview of task-related connectivity patterns in healthy subjects. Results are organized regarding frequency bands and age groups investigated. Upwards arrows (↑) indicate an increased connectivity during receptive (simple solid line) or expressive (dashed lines) language task, whereas downwards arrows (↓) indicate decreased connectivity.

et al., 2011; Kühn-Popp et al., 2016; Whedon et al., 2016) and grade-schoolers (Yang et al., 2005). Specifically, between 5 and 10 months of age, an increase in resting-state EEG coherence in the theta–alpha band (6–9 Hz) within left frontal regions seems to be associated with higher cognitive functioning, including receptive language at 3 years of age (Whedon et al., 2016). This association, however, might not be specific to language functions because the authors reported a mediating influence of the level of attentional control at the age of 2 years. Another study showed that, in the theta band (4–6 Hz), a pattern of less proximal (left-frontal to left-central) but more distal (left-frontal to left-occipital) resting state FC at 14 months old is negatively associated with the number of words expressed at the age of 2 years, as reported by the parents (lower vocabulary group; determined by the median split of the MacArthur Communicative Developmental Inventory (MCDI) results; Mundy et al., 2003). The same group also showed that at 18 months of age a ratio of higher proximal synchrony in the right hemisphere (right-frontal to right-central) is positively associated with vocabulary outcome (MCDI; total words) at 2 years old (Mundy et al., 2003).

At 14 months of age, a theta–alpha band (6–9 Hz), FC pattern of more proximal and less distal coherence appears to be specifically and positively associated with later language functioning, regardless of the child's IQ (Kühn-Popp et al., 2016). Accordingly, those results indicate that maturation of EEG coherence in the left hemisphere, established by the ratio of short-distance/long-distance connections, is positively correlated with preverbal communicative abilities at 15 months of age (e.g., pointing at objects) and with verbal communication skills at 48 months of age (epistemic language; Kühn-Popp et al., 2016). Congruently, left short-distance (parietotemporal) connectivity dominance in the theta band of preschoolers (32–64 months of age) during story listening shows exclusive positive correlation with language performance (no correlation with nonverbal cognitive performance or with chronological age), as assessed by the Kaufman Assessment Battery for Children

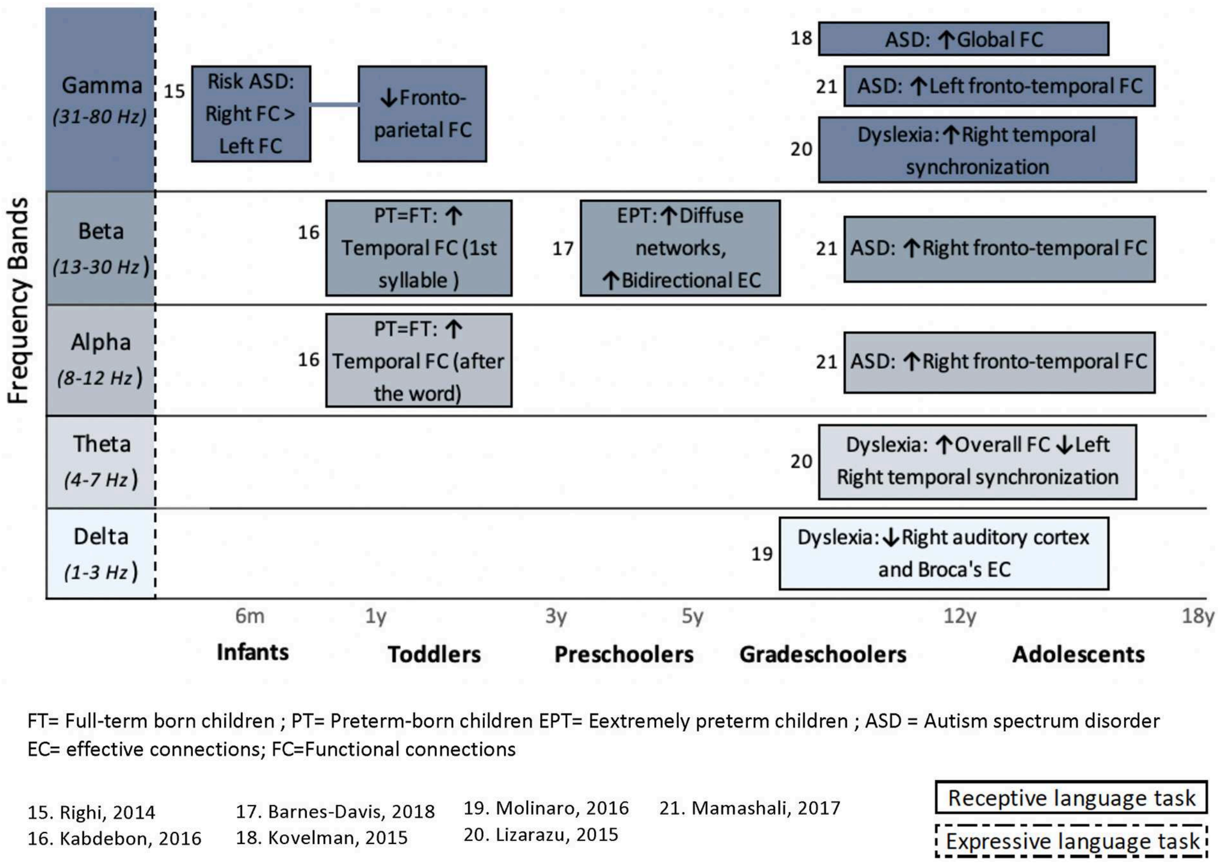


FIGURE 5 | Overview of task-related connectivity patterns in clinical populations compared to healthy subjects. Upper arrows (↑) indicate an increased connectivity during either receptive (simple solid line) or expressive (dashed lines) language task in this clinical population compared to healthy children, whereas downward arrow (↓) indicates decrease FC correlation in this clinical population compared to healthy children.

at the same age (Expressive Vocabulary and Riddles subtests; Kikuchi et al., 2011).

In older children (6–8 years old), participants with high language functioning (verbal IQ >110, as assessed by the Wechsler Intelligence Scale for Children III) had an increased chance of higher correlations between homologous hemispheric regions (homologous interhemispheric correlations), compared to those who were classified as having a low verbal functioning (verbal IQ <90; Yang et al., 2005). This was apparent in several regions (frontal, parietal) and mostly in the theta and alpha bands. In contrast, higher connectivity in interhemispheric central regions (delta and beta) was associated with lower language abilities.

However, one study reported non-significant correlations between coherence indices and language functioning. Marshall et al. (2008) highlighted environmental impacts on cerebral connectivity in young children, even though no significant correlation with language or cognitive functioning was found. They reported that EEG patterns in 42-month-old children placed in foster care before the age of 24 months differed from those of children placed in institutional care, the former showing lower short-distance connectivity. Specifically, in the foster-care group, intrahemispheric connections between frontal-central and frontal-temporal regions were characterized by lower connectivity in theta–alpha (6–10 Hz) and alpha–beta (11–18 Hz) bands. The authors did not link this difference to language abilities (no significant results) but instead to environmental conditions (foster care vs. institutional care).

Finally, an extensive longitudinal study including 508 children between 2 months and 16.5 years of age investigated developmental differences between sexes, using EEG coherence (Hanlon et al., 1999). However, no behavioral data were used to associate coherence patterns with language functioning. Results illustrated a sex difference in development, whereby girls presented earlier development of comprehensive language networks in theta neural networks than boys. Results also suggested that girls have more complex interconnection patterns between paired sites, particularly in those involving the temporal lobes.

Results from coherence in clinical population

Coherence for FC analyses was also used in several studies that included children with or at risk of neurodevelopmental conditions and therefore known to have vulnerable language functions. More specifically, included in this section are those studies using coherence as FC analyses and that focused on children with ASD, CHD, language learning impairment (LLI), stuttering, and dyslexia.

Children with CHD are known to be at higher risk of speech and language delays (Hövels-Gürich et al., 2008; Hövels-Gürich and Mccusker, 2016; Fourdain et al., 2019) It is in this context that Williams et al. (2012) investigated the predictive value of neonatal EEG frequency power analysis for later language development in children with CHD. Results revealed predictive value of the delta-relative power for language skills at 18 months of age, as assessed by the Bayley Scales of Infant Development (BSID). However, association between language functioning and coherence measures did not achieve significant results, despite the high correlation between BSID cognitive scores and beta's interhemispheric (left frontal polar to right frontal polar) and intrahemispheric (left frontal polar to left occipital) coherence. According to the authors, this may have been due to the small sample size ($n = 13$ participants).

Autism spectrum disorder is a neurodevelopmental disorder commonly associated with verbal and communicative dysfunctions (Mcdaniel et al., 2018). In three studies identified in this review, alteration of language task-related coherence was associated with ASD (Righi et al., 2014; Kovelman et al., 2015; Mamashli et al., 2017). However, no direct association was made with language functioning.

One publication aimed to identify an early electrophysiological biomarker for later ASD diagnosis (Righi et al., 2014). Electroencephalography recordings were performed for 6-month-old infants at high risk (HR, meaning siblings of children that were already diagnosed with ASD) and low-risk (LR) for ASD, done while listening to speech sounds. A higher right than left hemispheric coherence in the gamma band was observed in all children, with no difference between groups (HR vs. LR). At 12 months of age, analyses in LR and HR groups revealed no remaining hemispheric lateralization differences. Interestingly, HR infants showed significantly reduced task-related FC between frontal and parietal regions, compared to LR infants. Although these results must be replicated using a larger sample, this association seems to identify a potential 12-month predictive marker for clinical outcomes (Righi et al., 2014). These results also point out that genetic vulnerability for autism, that is, having a full sibling diagnosed with ASD, can potentially be assessed in the first year of life, based on differences in neural integration.

The two other published studies that used coherence involved older children with confirmed ASD diagnosis. Important differences were identified in FC patterns between healthy children and those diagnosed with ASD. Results of a preliminary study by Kovelman et al. (2015) indicated differences in cerebral coherence between ASD and control groups (8–12 years old) during a language task. In particular, EEG coherence measures during familiarization with a new language, including statistical learning for discrimination between adjacent syllables, were higher in children with ASD and had predictive value for ASD diagnosis. Coherence measures during the familiarization phase showed improved identification of ASD diagnosis, compared to coherence measure at rest, thus suggesting that language learning abilities are different in children with ASD, compared to typically developing (TD) peers.

Finally, Mamashli et al. (2017) used an MEG tonal mismatch paradigm in children (9–15 years old) with ASD. The MEG recording revealed an increase in frontotemporal coherence in the ASD group relative to the TD group, in response to both standard and deviant stimuli. This manifested in the gamma band for the left hemisphere and in the alpha and beta bands for the right hemisphere. When coherence was normalized with respect to the standard condition, the differences between groups were no longer significant. However, when the same stimuli were presented against a noisy background, the normalized coherence remained greater in ASD group, and this for the beta band in the left frontotemporal regions (not illustrated in **Figure 5**). According to the authors, this may suggest that, for ASD children, reduced speech comprehension in noisy surroundings is due to a lower involvement of frontal control mechanisms. These results imply that auditory processing, when done against a noisy background, results in altered functional networks in this group of patients.

Overall, studies in children with ASD demonstrated several distinct characteristics of functional neuronal networks associated with auditory and language processing, which are in line with typical difficulties in language functions associated with ASD. Knowing the characteristics of cerebral networks could potentially allow an early identification of children at higher risk of developing ASD.

Two studies involved participants with oral language disabilities, such as language disorder or childhood-onset fluency disorder (stuttering). Vasil'yeva and Shmalei (2013) were interested in brain coherence of male preschoolers (3–5-year-old boys) with neurosis-like stammering. These children showed generally stronger global coherence in delta and beta oscillations than did healthy children. Compared to healthy controls, theta band synchrony in interhemispheric frontal regions was also increased for the stammering group, although a smaller number of connections was observed in children who stutter than in healthy children. Finally, in all frequency bands, interhemispheric coherence was higher in preschoolers with neurosis-like stammering than in the control group. These results suggest that, in children with this kind of speech disturbance, the specialization of functions of the left and right hemispheres, as well as the interhemispheric asymmetry, is less expressed.

Finally, for children (6–11 years old) with non-verbal learning disorders, Njiokiktjien et al. (2001) reported a right lateralized decrease of intrahemispheric coherence, in contrast with children with LLI, who showed reversed lateralization. This difference was higher in the gamma band. Again, these M/EEG FC results suggest that hemispheric functional brain alterations are related to specific language development disorders.

Results From Phase Synchronization

Instead of investigating the relation between the amplitudes of the signals, one could also evaluate how the phases of the considered signals are coupled, the so-called phase synchronization measures. Among the many phase synchronization measures proposed in the literature, one of the most used is the PLV, which evaluates the phase difference between two signals (Lachaux et al., 1999). When two brain areas are functionally connected, the phases of their signals are assumed to evolve together; therefore, the difference in their phases should be constant (Bruña et al., 2018).

Results from phase synchronization in healthy children

Three studies combined phase synchronization metrics: two with an FC matrix (Doesburg et al., 2012; Youssofzadeh et al., 2017) and one with EC metrics (Barnes-Davis et al., 2018). Results from these three will be included in the sections on graph theoretical approaches and EC, respectively.

Two other studies drew on phase synchronization metrics (PLV) in healthy children: one in a mismatch paradigm (receptive task) and the other in an expressive language task.

At around 1 year of age, results during an audiovisual paradigm revealed an increased large-scale communication between brain regions in the mismatch condition (a heard sound does not match the previously presented symbol), compared to the match condition (sound and symbol match; Asano et al., 2015). This occurred in the alpha–beta band (12–15 Hz) and was more prominent in the left hemisphere. According to the authors, this indicates that audiovisual integration requires a greater effort in the mismatch condition (Asano et al., 2015).

In adolescents (17 years old), an expressive language task (verb generation) resulted in an increased gamma-band synchronization among task-activated cortical regions (Doesburg et al., 2012). Moreover, there was a theta modulation of interregional gamma synchrony between several pairs of activated brain regions, mostly in the left frontal cortex. This reflects the involvement of gamma-band synchronization in language production and the role of low-frequency rhythms (theta), which modulate high-frequency connectivity in adolescents.

Results from phase synchronization in clinical population

One study used phase synchronization metrics (PLV) in task-related paradigms, in a vulnerable population, namely, children born prematurely. In fact, several studies report impairments of cognitive and behavioral functions, including language abilities, related to premature birth (weeks of gestation ≤ 37; e.g., Aarnoudse-Moens et al., 2009; de Kieviet et al., 2012). In our sample, one study used PLV for FC analyses in prematurely born children (27–30 weeks of gestation). Kabdebon et al. (2015) compared spatial synchrony and phase coincidence of EEG oscillations during syllabic learning in 8-month-old preterm-born and term-born children (corrected age for preterm-born). They did not find any differences between groups, suggesting similar language processing at 8 months of age. In both groups, an increase in the PLV was observed first in the beta band (13–18 Hz; during the first syllable) and later in alpha (8–12 Hz; after the word) over the left and right temporal areas (Kabdebon et al., 2015).

Using auditory stimuli in children (8–14 years old) and adults with dyslexia, another study found that, compared to a control group, dyslexic participants presented stronger synchronization and an absence of right hemispheric neural synchronization, related to low frequency (4 Hz; Lizarazu et al., 2015). On the other hand, for high frequencies (30 Hz), adults but mainly children with dyslexia show a rightward, instead of bilateral hemispheric lateralization. According to the authors, this may suggest that speech processing in dyslexic children relies more heavily on syllabic-rate information, compared to skilled reader peers.

Results From Network Analysis

Graph theory analysis looks at the brain as a complex network consisting of a collection of nodes connected by edges, in order to comprehend the topological organization of brain networks (Tahmasian et al., 2015).

Results from network analysis in healthy children

Two studies applied graph theoretical analysis into MEG results to investigate the organization of expressive language networks, from preschool age to adolescence (4–18 years old). Even though both used a verb generation task during MEG, and derived networks from phase synchronization metrics, their conclusions were not identical.

In the first of the two, results from a verb generation task revealed a developmental shift of the beta band lateralization in language production when children (4–6 years old) were compared to adolescents (16–18 years old): hubs were most lateralized in adolescents, whereas younger children showed a more bilateral distribution, or even a right-hemispheric pattern (Youssofzadeh et al., 2017).

The second study showed that connectivity within language-related areas (left angular gyrus, left precentral gyrus, right inferior orbital gyrus, and right rolandic operculum) increased with age (Doesburg et al., 2016). This was true for language production in the theta band. Increased FC during an expressive language task was also observed in higher frequency bands (alpha and beta). However, this increase was primarily found in brain areas associated with visual processing and thus might rather be associated with processing of the stimulus than to language-related task demands. Developmental analysis suggested significant differences between age groups: larger connectivity networks in adolescents (14–18 years old), compared to younger children (4–9 years old), and a stronger task-dependent increase of connectivity (expressed as theta coherence) in language-related areas, especially in frontal regions. Finally, theta-band connectivity measures showed a significant association with verbal language functioning (assessed with the Peabody Picture Vocabulary Test and the Expressive Vocabulary Test). Thus, the strength of task-dependent network connectivity was associated not only with a maturational pattern but also with language abilities (Doesburg et al., 2016).

Results from graph theoretical analysis in clinical population

Zare et al. (2016) developed a machine learning approach based on EEG network characteristics (efficiency and leaf number) in 6-month-old infants. They aimed at determining, based on family history, the risk of LLDs. Relying on functional connectivity measures, this work allowed for the accurate stratification of the children into low-risk (LR) and high-risk (HR) groups for LLD. Early brain networks revealed a reduced cortical communication capacity in HR infants, showing a network that was both decentralized (as revealed by the clustering index in the delta and alpha) and less efficient (as revealed by a decreased efficiency in the delta, theta, and alpha). Based on complex EEG patterns with support vector machine, it was possible to classify the children into HR and LR groups with approximately 80% accuracy (specificity of 89% and sensitivity of 92%).

Directionality of Language Networks (Effective Connectivity)

Effective connectivity reveals the directionality of information flow in particular brain regions and the causal and dynamic influences of one region on another (Stephan and Friston, 2010; Friston, 2011). Two methods of EC were used in the studies selected for review: partial directed coherence, a frequency-domain representation of the concept of Granger causality (Baccalá and Sameshima, 2001) and the PSI, a method based on phase differences in signals over a specified frequency range (Nolte et al., 2008).

Effective connectivity in healthy children

Only one study used EC metrics to study language networks in healthy children during an expressive language task. Kadis et al. (2016) reported an increased number of effective connections (PSI) with age, between 5 and 18 years. More importantly, different task-related EC patterns seemed to emerge among frequency bands. Analysis of lower frequency bands revealed more local, rostrally directed connectivity patterns in the left frontal region. At higher frequencies, EC increasingly involved distal and interhemispheric nodes. In alpha and gamma, bidirectional information transfer was observed between left and right frontal and posterior temporal nodes, whereas in the gamma band, the right posterior temporal region emerged as an important driver of Wernicke (left posterior temporal) and Broca (left frontal) regions.

Effective connectivity in clinical population

Phase slope index was also used to compare EC (PSI) and FC patterns (PLI) between extremely prematurely born children (EPT; <28 weeks of gestation) and their term-born (TB) peers [37–42 weeks of gestation; (Barnes-Davis et al., 2018)]. At preschool age (4–6 years old), bilateral functional networks, including temporal and parietal regions, were revealed in both EPT and TB children during a receptive language task. On the other hand, the beta band indicated increased FC in language networks, as well as a more diffused network in EPT children, compared to TB. Moreover, analysis of EC suggested more bidirectional connections in EPT within bitemporal areas of the network, compared to TB, where fewer bidirectional networks or more unidirectional networks were identified. Effective connectivity analysis also revealed that hyperconnectivity patterns in EPT were attributable to a greater information flux drive from the right hemisphere. Nevertheless, because those differences in connectivity patterns were not correlated with language performance, it was reported to be an effect of the clinical condition only (i.e., prematurity). Consequently, the authors assumed that their findings indicated an efficient reorganization of cerebral language networks, allowing the maintenance of language abilities in EPT children (Barnes-Davis et al., 2018).

Neuronal response while listening to low-frequency speech (<10 Hz), in grade-schoolers (8–14 years old) with dyslexia, was overall less synchronized, compared to normal readers (Molinaro et al., 2016). More specifically, during language stimulation (meaningful sentences), reduced delta synchronization and impaired feed forward functional coupling (partial directed coherence) were found between the right auditory cortex and the left inferior frontal gyrus.

DISCUSSION

We systematically reviewed 24 studies that assessed M/EEG functional networks associated with language in children. The great variability in study populations, sample size, and methodology precluded us from conducting a meta-analysis. Instead, we synthesized and critically appraised findings on the use of functional or EC in the study of spoken language in children.

Summary of the Main Observations

In order to characterize functional networks involved in language development, first considered were results reported in 13 articles on the study of TD children, and which used FC and EC analyses. The findings of most of the reviewed studies suggested that theta neural oscillations play a crucial role in healthy language development. In the theta band, a greater left resting-state coherence in early childhood seems to be associated with higher language functioning, either at the time of M/EEG recording (Kikuchi et al., 2011) or at a later age (Mundy et al., 2003; Kühn-Popp et al., 2016; Whedon et al., 2016). In older children (grade-schoolers to adolescents), associations between connectivity patterns and language abilities are not found only in theta, but in most frequency bands (delta, theta, alpha, and beta). The differences in frequency bands in relation to age might reflect typical brain maturation. Indeed, cerebral maturation in children has been associated with a global decrease of slow-wave activity, including theta oscillations, and an increase in higher frequencies (Uhlhaas et al., 2010). Thus, even though theta-band connectivity shows significant correlation with language abilities at all ages (**Figure 3**), it is critical to look at all different frequency bands, especially in older children (grade-schoolers and adolescents).

Further, theta frequency band has been related to syllabic processing (Giraud and Poeppel, 2012; Meyer, 2018), and increases in theta activation have been found for tasks that include verbal working memory (Friederici and Singer, 2015; Meyer, 2018). Syllabic processing of human language constitutes one of the fundamental stages of bottom-up language processing, and there is evidence that it is established *in utero*, before term age (Mahmoudzadeh et al., 2013; Skeide and Friederici, 2016). The predictive value of theta coherence for early language comprehension in infants may thus be explained by the fundamental role of syllabic processing in later language acquisition. Given the assumed relation between theta band coherence and working memory, studies addressing language networks should also apply language paradigms that allow for the differentiation between higher-order cognitive functions and different stages of language processing.

The investigation of FC or EC networks using a language task during M/EEG recording reveals results distributed across all frequency bands. The involvement of the various frequency bands probably varies based on the nature of the task (e.g., active lexical discrimination vs. passive oddball paradigm), the language modality (expressive vs. receptive), and the level of language processing (e.g., syllabic vs. semantic). That being said, results from EC patterns in expressive language paradigm vary considerably depending on the frequency bands (Kadis et al., 2016). An age-related increase is shown in left effective connections, whereas higher frequencies reveal more bilateral effective connections with increasing age (Kadis et al., 2016).

For healthy children, the majority of studies using task-dependent connectivity analysis reveal increased left FC during receptive (Kikuchi et al., 2011; Asano et al., 2015) and expressive (Doesburg et al., 2012, 2016; Youssofzadeh et al., 2017) language paradigms. This occurs as early as 11 months of age (Asano et al., 2015) and appears to be constant throughout development. Interestingly, when it comes to examining the pattern of task-related FC in populations at risk of language disorders, in comparison with neurotypical children, differences are prominently characterized by a tendency for greater FC in the right hemisphere (Righi et al., 2014; Lizarazu et al., 2015; Mamashli et al., 2017).

Results from studies targeting clinical populations, mainly children at high risk of or suffering from language disabilities, also contribute to the understanding of the interactions between language abilities and the brain regions associated with language acquisition. In this review, we included 11 studies that addressed FC and EC patterns of language networks in different clinical populations. In children with speech disturbances (language learning disorders or stuttering), the functional specialization in the left and right hemispheres and the interhemispheric asymmetry typically seen in language networks seem altered (less hemispheric asymmetry observed). However, in populations at risk of language disabilities, such as ASD, preterm children, and infants with CHD, there are no clear or replicable FC profiles associated with language functioning that arise from the current literature. Although differences are observable between clinical and control groups, they seem to be more attributed to the signature of the underlying clinical condition, rather than to language functioning itself. More studies are needed to better understand the brain substrates of language alterations and vulnerabilities in these populations.

These results are consistent with the conclusion from Weiss-Croft and Baldeweg (2015), who found that left language lateralization was well established by the age of 5 years. However, our results suggest that, before the first birthday, left lateralization is already apparent when a receptive language paradigm is performed (Asano et al., 2015). Moreover, a greater left connectivity before 5 years of age has been correlated with better language abilities (Mundy et al., 2003; Kikuchi et al., 2011; Kühn-Popp et al., 2016; Whedon et al., 2016). Thus, M/EEG research points toward an earlier implementation of left lateralization in language networks than was concluded from studies done with fMRI. This is probably due to the suitability of electrophysiological techniques for studying young children. Furthermore, the impaired left lateralization in populations at risk of language impairments attests to the importance of the early development of left functional networks (Righi et al., 2014; Barnes-Davis et al., 2018) and its maintenance in later development (Lizarazu et al., 2015; Mamashli et al., 2017).

The developmental trajectory of FC of language networks evolves significantly with age, with the presence of greater connectivity networks in adolescents, compared to younger children (Doesburg et al., 2016; Kadis et al., 2016; Poblano et al., 2016; Youssofzadeh et al., 2017), but also more local and less bilateral networks as age increases (Kikuchi et al., 2011; Doesburg et al., 2016; Kadis et al., 2016). In line with findings of fMRI studies, strong local networks may actually reflect both processes related to cerebral specialization and automatized language processing, which require less top-down regulation and thus involves fewer network interactions (Weiss-Croft and Baldeweg, 2015).

Nonetheless, the exact timeline of maturational processes in language networks is not yet fully understood. This may be due in part to the great intervariability of typical development. Also, many studies included only a limited age range or did not have sufficient participants per age group to permit reliable conclusions regarding developmental changes. The importance of accounting for age-related changes has previously been emphasized in fMRI studies, in order to correctly interpret associations between network characteristics and language capacities (e.g., Weiss-Croft and Baldeweg, 2015; Rimmele et al., 2018). On the other hand, the methodological heterogeneity (e.g., language paradigms, cognitive assessments, connectivity algorithms) between developmental studies on brain correlates of language processing do not allow the drawing of a clear maturational timeline.

Finally, one should consider that sex differences may impact the development of FC patterns, as stated by Hanlon et al. (1999). In fact, the importance of integrating sex analysis in research is now well-established (Tannenbaum et al., 2019), and the sex differences of brain development have been documented (Gur and Gur, 2016, 2017; Kaczkurkin et al., 2019). In a recent systematic review, Etchell et al. (2018) highlighted sex differences in brain language structure and function. However, they concluded that these differences do not necessarily lead to differences in language task performance. It is therefore possible that boys and girls employ different but equally effective cognitive strategies for certain tasks, which leads to minor differences in performance as evidenced by brain function but not in the behavioral performance itself. Consequently, it is important that subsequent studies consider possible sex differences when characterizing language networks.

A better understanding of the association between language functions and the different characteristics of brain networks should include normal variation patterns that are not related to language difficulties. Understanding the normal development of functional language networks would enable earlier identification of children at risk of language difficulties. Currently, language impairment is often detected only at an age at which evidence of healthy language functions can be formally assessed (Prelock et al., 2008). When a pathology is present, however, it could be crucial to initiate early intervention in order to support language development and increase quality of life for these children.

Methodological Considerations

This review shines light on the heterogeneity of methodological approaches used in the study of language functions in children, through the use of FC and EC. Beyond the neuroimaging method used (EEG vs. MEG), the type of analyses and their nomenclature vary greatly between research groups. Functional brain connectivity and EC analyses are indeed still recent, and to date, there is no consensus on which methods are to be advocated, highlighting the importance of summarizing the current state of knowledge and pursuing further research in this field. This would not only describe the various methods available, but also assess their respective pros and cons, in order to select the appropriate technique for specific experimental conditions and samples. This will ultimately support the production of more reliable and robust results and provide clear directions for future studies. Methodological heterogeneity is not only an issue in EEG and MEG, but also poses an obstacle to reliable conclusions about language networks estimated with other neuroimaging techniques, such as fMRI (Weiss-Croft and Baldeweg, 2015), hence the need to establish common standards of best practice.

Nevertheless, the number of M/EEG studies identified indicates that coherence and phase-locking measures may have high utility in language research, because these metrics were used in the majority of the published articles in the domain. These approaches achieved popularity because of their simple algorithms and fast computation. However, although coherence has been the most widely used FC method in this field, this does not necessarily mean it is the preferred method, nor the most fruitful. In fact, coherence may cause false-positive results, due to source leakage between local regions (Brookes et al., 2014; Kida et al., 2015). To overcome these challenges, many algorithms have been developed in the last few years. The Imaginary Part of Coherency (Nolte et al., 2008) and PLI are metrics that are less affected by the influence of common sources and active reference electrodes. They were introduced to facilitate the estimation of phase synchronization but have not been used much in the research of language development (none for Imaginary Part of Coherency and twice for PLI). Yet, the simplest method for reducing the influence of leakage on the estimation of connectivity is a leakage-invariant metric (O'reilly et al., 2017).

Conversely, the use of task-evoked EC metrics such as Granger causality and PLI in this context is recent and remains limited, given that only three research teams have applied them since 2016. Thus, little is known about the directionality (EC) of oral language networks in children.

To date, the use of EEG is more frequent than MEG for the investigation of language-related brain connectivity in children (14 and 10 articles, respectively), certainly because of the higher accessibility, lower cost, and ease of use of the EEG technique.

Methodological Limitations of Reviewed Studies

The primary methodological limitation of most studies reviewed was the failure to directly examine the association between brain FC patterns and objective language skills as assessed by standardized behavioral tests. In addition, in those studies that did evaluate language abilities, assessment of overall cognitive functioning was not always performed. Thus, the observed disturbance could indicate a lower global cognitive functioning, rather than a specific effect of language difficulties. A clear distinction between language and global cognitive functioning is therefore critical when investigating links between connectivity patterns and language performance. Relationships between brain activity and behavior must be addressed, especially in the context of clinical populations, where the disturbance in FC patterns associated with the neurodevelopmental condition must be distinguished from the disturbance specific to language functions alterations. For instance, in contrast to healthy children, M/EEG FC differences in children with CHD or born prematurely are not always associated with actual differences in language skills. The lack of attention to these relationships may be partially explained by the small sample sizes of the studies, which led to poor statistical power.

Finally, the results from various studies emphasized the difficulty of applying FC analysis derived from M/EEG data. Source localization of cerebral activity, captured on the surface of the scalp, represents a particular challenge for sensor-space analysis. This is known as the inverse problem, which may lead to inaccurate identification of cerebral networks (e.g., Nunez et al., 1997; Sakkalis, 2011; Barzegaran and Knyazeva, 2017; Abreu et al., 2018, 2019). Also, the effect of volume conduction, which is a mix of several signals within one sensor, and which originate from identical cerebral regions, makes critical a direct derivative from sensors to cerebral representation. Source-space analysis tries to overcome this downside and uses models that aim for a more accurate reconstruction of the true sources of the signal (Schoffelen and Gross, 2009). The conduction of source analyses seems particularly important when one is aiming to interpret FC, because the same cerebral activation is measured with different sensors and may potentially result in false conclusions regarding connected regions. Recently, it has been shown that source-space analyses seem accurate mostly when using high-density EEG, but result in limited interpretation of the more common low-density EEG (Barzegaran and Knyazeva, 2017). Also, some of the approaches to source analysis require certain assumptions be made about the underlying network, which may not be accurate for all data sets (Daunizeau and Friston, 2007). In particular, in children (where networks are developing) or in clinical populations (where networks may be altered), it can be risky to assume a certain network composition. These limitations need to be taken into consideration when interpreting some of the findings on functional networks that are reported in this review. While studies that applied sensor-space analysis may overestimate functional connectivity, the interpretation of findings based on source-space analysis, especially in low-density EEG, may be less susceptible to this same overestimation. Finally, some studies might not have verified specific assumptions for their source-model, which limits their interpretation. This issue may occur especially in studies that include clinical populations, where characteristics of cerebral activation may be altered.

General Utility of M/EEG Connectivity Analysis

By providing information about temporal coupling between cortical areas (milliseconds time scale) and frequency bands of neural oscillations, both MEG and EEG are well-suited to study the development of language networks. They offer a quiet testing environment, which facilitates the use of language tasks. Moreover, they provide excellent temporal resolution, allowing analyses that target an immediate response to specific tasks or stimuli.

Because EEG is less sensitive to movement than other techniques (e.g., fMRI), thus allowing a certain mobility and tolerating articulatory movements, it is highly relevant for language assessment in pediatric populations. Furthermore, the low cost of EEG justifies its use for the investigation of developmental trajectories, which requires longitudinal design with multiple recordings over time. On the other hand, spatial and temporal data available from MEG allow the investigator to track both the neural timing and location associated with language and thus to efficiently map the trajectories of language networks. Regardless of the neuroimaging technique employed, the use of FC is highly relevant in research on children, because it allows acquisition at rest, without requiring that a task be performed, as it is in traditional ERP paradigms. Furthermore, the length of time required for data acquisition can usually be shorter, compared to task paradigms. Finally, a better understanding of FC M/EEG analysis and an evaluation of their usefulness are essential for future research and for the potential use of these techniques in clinical contexts.

Limits of This Review

Although this systematic review goes beyond a simple revision of the literature, it does not include any statistical analysis of the reviewed studies, as would have provided a meta-analysis. The reader should therefore take into account the fact that the current findings represent qualitative and not quantitative results. The methodological heterogeneity of the included studies, with respect to their paradigms, the types of FC and EC analysis, as well as the large age range of the children investigated, is in itself a limitation for the generalization and integration of the results.

Compared to other neuroimaging techniques, both MEG and EEG stand out because of their high temporal resolution. This is of particular importance in language paradigms, where tonal differences occur at a fast rate. However, both methods have a relatively low spatial resolution, which leads to a rather large-scale localization of cerebral activity when compared to techniques such as fMRI. Thus, the present findings about functional language brain networks permit only limited spatial interpretation.

Finally, given that we mainly reviewed studies that considered FC as a measure of neuronal networks, we would like to acknowledge that FC bears an index of statistical dependency. More precisely, it allows the estimation of the correlation between cerebral activation, measured simultaneously with different electrodes or sensors located over different cerebral locations. Thus, it does not allow causal conclusions about brain networks. Only three studies (Kadis et al., 2016; Molinaro et al., 2016; Barnes-Davis et al., 2018) included EC analysis that allowed causal conclusions about interactions within functional language networks. Future studies should definitely include EC analysis that allows for more advanced characterization of cerebral language networks.

CONCLUSION AND FUTURE DIRECTIONS

The analysis of brain functional connectivity and EC through the use of M/EEG data is a common emphasis of ongoing developmental research, but many unanswered questions remain regarding the brain correlates of language development. To our knowledge, this is the first systematic review to summarize the current state of knowledge on linguistic electrophysiological patterns of brain connectivity in the pediatric population. It provides a detailed portrait of the relevant MEG and EEG data analysis methods that have been used in that context. Future research should consider the different FC analyses available, in order to choose the appropriate tools and paradigms. Overall, the results of the reviewed studies are highly heterogeneous, precluding the possibility of drawing clear and quantitative conclusions and showing the importance of pursuing research in this field. Future work will enlighten on the brain substrates of language development and may also have important clinical impacts, for example, leading to the identification of early neuroimaging markers associated with altered language development in populations at high risk of language disabilities. It would also allow the identification of children at higher risk of language difficulties, in order to provide early and individualized intervention (Jeste et al., 2015). However, studies with significantly larger sample sizes, as well-normative data, are needed in order to be able to use these tools in a clinical context.

DATA AVAILABILITY STATEMENT

Datasets are available upon request to the corresponding author.

AUTHOR CONTRIBUTIONS

IG developed search procedure, performed database searches and reviewed articles for inclusion/exclusion based on the title, and abstract. IG and AH reviewed articles for inclusion/exclusion based on the full text, extracted data from articles, analyzed the extracted data, and wrote the manuscript. AG supervised all aspects of the systematic literature review, preparation of the manuscript, revision, editing, and final intellectual content. PV contributed to intellectual content and provided comments on the manuscript.

FUNDING

This study was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2015-04199, 2015–2020), a Project Grant from the Candian Institutes of Health Research (CIHR, #419351, 2019–2024), and a Canada Research Chair (CRC, #950-232651) hold by AG.

REFERENCES

Aarnoudse-Moens, C. S., Weisglas-Kuperus, N., Van Goudoever, J. B., and Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics* 124, 717–728. doi: 10.1542/peds.2008-2816

Abreu, R., Leal, A., and Figueiredo, P. (2018). EEG-informed FMRI: a review of data analysis methods. *Front. Hum. Neurosci.* 12:29 doi: 10.3389/fnhum.2018.00029

Abreu, R., Leal, A., and Figueiredo, P. (2019). Identification of epileptic brain states by dynamic functional connectivity analysis of simultaneous EEG-fMRI: a dictionary learning approach. *Sci. Rep.* 9:638 doi: 10.1038/s41598-018-36976-y

Ardila, A., Bernal, B., and Rosselli, M. (2015). How localized are language brain areas? a review of brodmann areas involvement in oral language. *Arch. Clin. Neuropsychol.* 31, 112–122. doi: 10.1093/arclin/acv081

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., and Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex* 63, 196–205. doi: 10.1016/j.cortex.2014.08.025

Aydore, S., Pantazis, D., and Leahy, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage* 74, 231–244. doi: 10.1016/j.neuroimage.2013.02.008

Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990

Barnes-Davis, M. E., Merhar, S. L., Holland, S. K., and Kadis, D. S. (2018). Extremely preterm children exhibit increased interhemispheric connectivity for language: findings from fMRI-constrained MEG analysis. *Dev. Sci.* 21:e12669. doi: 10.1111/desc.12669

Barzegaran, E., and Knyazeva, M. G. (2017). Functional connectivity analysis in EEG source space: the choice of method. *PLoS ONE* 12:e0181105. doi: 10.1371/journal.pone.0181105

Bastos, A. M., and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9:175. doi: 10.3389/fnsys.2015.00175

Berwick, R. C., Friederici, A. D., Chomsky, N., and Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17, 89–98. doi: 10.1016/j.tics.2012.12.002

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Boersma, M., Smit, D. J., Boomsma, D. I., De Geus, E. J., Delemarre-Van De Waal, H. A., and Stam, C. J. (2013). Growing trees in child brains: graph theoretical analysis of electroencephalography-derived minimum spanning tree in 5- and 7-year-old children reflects brain maturation. *Brain Connect.* 3, 50–60. doi: 10.1089/brain.2012.0106

Bowyer, S. M. (2016). Coherence a measure of the brain networks: past and present. *Neuropsychiatr. Electrophysiol.* 2:1–12. doi: 10.1186/s40810-015-0015-7

Bressler, S. L., and Seth, A. K. (2011). Wiener-Granger causality: a well established methodology. *Neuroimage* 58, 323–329. doi: 10.1016/j.neuroimage.2010.02.059

Brookes, M. J., Woolrich, M. W., and Price, D. (2014). "An introduction to MEG connectivity measurements," in *Magnetoencephalography: From Signals to Dynamic Cortical Networks*, eds. S. Supek and C.J. Aine (Berlin: Springer), 321–358.

Bruña, R., Maestú, F., and Pereda, E. (2018). Phase locking value revisited: teaching new tricks to an old dog. *J. Neural Eng.* 15:056011. doi: 10.1088/1741-2552/aacfe4

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–98. doi: 10.1038/nrn2575

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (2007). *Handbook of Psychophysiology.* New York, NY: Cambridge University Press.

Daunizeau, J., and Friston, K. J. (2007). A mesostate-space model for EEG and MEG. *Neuroimage* 38, 67–81. doi: 10.1016/j.neuroimage.2007.06.034

de Kieviet, J. F., Zoetebier, L., Van Elburg, R. M., Vermeulen, R. J., and Oosterlaan, J. (2012). Brain development of very preterm and very low-birthweight children in childhood and adolescence: a meta-analysis. *Dev. Med. Child Neurol.* 54, 313–323. doi: 10.1111/j.1469-8749.2011.04216.x

Doesburg, S. M., Tingling, K., Macdonald, M. J., and Pang, E. W. (2016). Development of network synchronization predicts language abilities. *J. Cogn. Neurosci.* 28, 55–68. doi: 10.1162/jocn_a_00879

Doesburg, S. M., Vinette, S. A., Cheung, M. J., and Pang, E. W. (2012). Theta-modulated gamma-band synchronization among activated regions during a verb generation task. *Front. Psychol.* 3:195 doi: 10.3389/fpsyg.2012.00195

Etchell, A., Adhikari, A., Weinberg, L. S., Choo, A. L., Garnett, E. O., Chow, H. M., et al. (2018). A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia* 114, 19–31. doi: 10.1016/j.neuropsychologia.2018.04.011

Fourdain, S., St-Denis, A., Harvey, J., Birca, A., Carmant, L., Gallagher, A., et al. (2019). Language development in children with congenital heart disease aged 12–24 months. *Eur. J. Paediatr. Neurol.* 23, 491–499. doi: 10.1016/j.ejpn.2019.03.002

Friederici, A. D., and Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci.* 19, 329–338. doi: 10.1016/j.tics.2015.03.012

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time Series. *J. Am. Stat. Assoc.* 77, 304–313. doi: 10.1080/01621459.1982.10477803

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063

Gur, R. C., and Gur, R. E. (2017). Complementarity of sex differences in brain and behavior: From laterality to multimodal neuroimaging. *J. Neurosci. Res.* 95, 189–199. doi: 10.1002/jnr.23830

Gur, R. E., and Gur, R. C. (2016). Sex differences in brain and behavior in adolescence: findings from the Philadelphia neurodevelopmental cohort. *Neurosci. Biobehav. Rev.* 70, 159–170. doi: 10.1016/j.neubiorev.2016.07.035

Hanlon, H. W., Thatcher, R. W., and Cline, M. J. (1999). Gender differences in the development of EEG coherence in normal children. *Dev. Neuropsychol.* 16, 479–506. doi: 10.1207/S15326942DN1603_27

Hassan, M., and Wendling, F. (2018). Electroencephalography source connectivity: aiming for high resolution of brain networks in time and space. *IEEE Signal Process. Mag.* 35, 81–96. doi: 10.1109/MSP.2017.2777518

Herrmann, C. S., Strüber, D., Helfrich, R. F., and Engel, A. K. (2016). EEG oscillations: from correlation to causality. *Int. J. Psychophysiol.* 103, 12–21. doi: 10.1016/j.ijpsycho.2015.02.003

Hövels-Gürich, H. H., Bauer, S. B., Schnitker, R., Willmes-Von Hinckeldey, K., Messmer, B. J., Seghaye, M. C., et al. (2008). Long-term outcome of speech and language in children after corrective surgery for cyanotic or acyanotic cardiac defects in infancy. *Eur. J. Paediatr. Neurol.* 12, 378–386. doi: 10.1016/j.ejpn.2007.10.004

Hövels-Gürich, H. H., and Mccusker, C. (2016). "Neurodevelopmental patterns in congenital heart disease across childhood: longitudinal studies from europe," in *Congenital Heart Disease and Neurodevelopment* (San Diego, CA: Academic Press), 41–53.

Jeste, S. S., Frohlich, J., and Loo, S. K. (2015). Electrophysiological biomarkers of diagnosis and outcome in neurodevelopmental disorders. *Curr. Opin. Neurol.* 28, 110–116. doi: 10.1097/WCO.0000000000000181

Kabdebon, C., Pena, M., Buiatti, M., and Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain Lang.* 148, 25–36. doi: 10.1016/j.bandl.2015.03.005

Kaczkurkin, A. N., Raznahan, A., and Satterthwaite, T. D. (2019). Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology* 44, 71–85. doi: 10.1038/s41386-018-0111-z

Kadis, D. S., Dimitrijevic, A., Toro-Serey, C. A., Smith, M. L., and Holland, S. K. (2016). Characterizing information flux within the distributed pediatric expressive language network: a core region mapped through fMRI-constrained MEG effective connectivity analyses. *Brain Connect.* 6, 76–83. doi: 10.1089/brain.2015.0374

Kida, T., Tanaka, E., and Kakigi, R. (2015). Multi-Dimensional dynamics of human electromagnetic brain activity. *Front. Hum. Neurosci.* 9:713. doi: 10.3389/fnhum.2015.00713

Kikuchi, M., Shitamichi, K., Yoshimura, Y., Ueno, S., Remijn, G. B., Hirosawa, T., et al. (2011). Lateralized theta wave connectivity and language performance in 2- to 5-year-old children. *J. Neurosci.* 31, 14984–14988. doi: 10.1523/JNEUROSCI.2785-11.2011

Kösem, A., and Van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Lang. Cogn. Neurosci.* 32, 536–544. doi: 10.1080/23273798.2016.1238495

Kovelman, I., Wagley, N., Hay, J. S., Ugolini, M., Bowyer, S. M., Lajiness-O'neill, R., et al. (2015). Multimodal imaging of temporal processing in typical and atypical language development. *Ann. N. Y. Acad. Sci.* 1337, 7–15. doi: 10.1111/nyas.12688

Kühn-Popp, N., Kristen, S., Paulus, M., Meinhardt, J., and Sodian, B. (2016). Left hemisphere EEG coherence in infancy predicts infant declarative pointing and preschool epistemic language. *Soc. Neurosci.* 11, 49–59. doi: 10.1080/17470919.2015.1024887

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c

Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P. M., Lerma-Usabiaga, G., et al. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Hum. Brain Mapp.* 36, 4986–5002. doi: 10.1002/hbm.22986

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., et al. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4846–4851. doi: 10.1073/pnas.1212220110

Mamashli, F., Khan, S., Bharadwaj, H., Michmizos, K., Ganesan, S., Garel, K. A., et al. (2017). Auditory processing in noise is associated with complex patterns of disrupted functional connectivity in autism spectrum disorder. *Autism Res.* 10, 631–647. doi: 10.1002/aur.1714

Marshall, P. J., Reeb, B. C., Fox, N. A., Nelson, C. A., and Zeanah, C. H. (2008). Effects of early intervention on EEG power and coherence in previously institutionalized children in Romania. *Dev. Psychopathol.* 20, 861–880. doi: 10.1017/S0954579408000412

Mcdaniel, J., D'ambrose Slaboch, K., and Yoder, P. (2018). A meta-analysis of the association between vocalizations and expressive language in children with autism spectrum disorder. *Res. Dev. Disabil.* 72, 202–213. doi: 10.1016/j.ridd.2017.11.010

Meng, L., and Xiang, J. (2016). Frequency specific patterns of resting-state networks development from childhood to adolescence: a magnetoencephalography study. *Brain Dev.* 38, 893–902. doi: 10.1016/j.braindev.2016.05.004

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur. J. Neurosci.* 48, 2609–2621. doi: 10.1111/ejn.13748

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLos Med.* 6, e1000097. doi: 10.1371/journal.pmed.1000097

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., and Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Hum. Brain Mapp.* 37, 2767–2783. doi: 10.1002/hbm.23206

Mundy, P., Fox, N., and Card, J. (2003). EEG coherence, joint attention and language development in the second year. *Dev. Sci.* 6, 48–54. doi: 10.1111/1467-7687.00253

Njiokiktjien, C., De Rijke, W., and Jonkman, E. J. (2001). Children with Nonverbal learning disabilities (NLD): coherence values in the resting state may reflect hypofunctional long distance connections in the right hemisphere. *Fiziol. Cheloveka* 27, 17–22. doi: 10.1023/A:1012335223507

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., et al. (2008). Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100:234101. doi: 10.1103/PhysRevLett.100.234101

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., et al. (1997). EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* 103, 499–515. doi: 10.1016/s0013-4694(97)00066-7

Oldham, S., and Fornito, A. (2018). The development of brain network hubs. *Dev. Cogn. Neurosci.* 36:100607. doi: 10.31234/osf.io/85u7z

O'reilly, C., Lewis, J. D., and Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLoS ONE* 12:e0175870. doi: 10.1016/j.dcn.2018.12.005

Palva, J. M., and Palva, S. (2018). Functional integration across oscillation frequencies by cross-frequency phase synchronization. *Eur. J. Neurosci.* 48, 2399–2406. doi: 10.1111/ejn.13767

Poblano, A., Castro-Sierra, E., Arteaga, C., and Pérez-Ruiz, S. J. (2016). Lexical tonal discrimination in Zapotec children. A study of the theta rhythm. *Bol. Med. Hosp. Infant. Mex.* 73, 84–89. doi: 10.1016/j.bmhimx.2015.09.010

Prelock, P. A., Hutchins, T., and Glascoe, F. P. (2008). Speech-language impairment: how to identify the most common and least diagnosed disability of childhood. *Medscape. J. Med.* 10, 136–136.

Righi, G., Tierney, A. L., Tager-Flusberg, H., and Nelson, C. A. (2014). Functional connectivity in the first year of life in infants at risk for autism spectrum disorder: an EEG study. *PLoS ONE* 9:e105176. doi: 10.1371/journal.pone.0105176

Rimmele, J. M., Gross, J., Molholm, S., and Keitel, A. (2018). Editorial: brain oscillations in human communication. *Front. Hum. Neurosci.* 12:39. doi: 10.3389/fnhum.2018.00039

Saby, J. N., and Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Dev. Neuropsychol.* 37, 253–273. doi: 10.1080/87565641.2011.614663

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Comput. Biol. Med.* 41, 1110–1117. doi: 10.1016/j.compbiomed.2011.06.020

Schoffelen, J. M., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Hum. Brain Mapp.* 30, 1857–1865. doi: 10.1002/hbm.20745

Skeide, M. A., and Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* 17, 323–332. doi: 10.1038/nrn.2016.23

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425. doi: 10.1016/j.tics.2004.07.008

Stam, C. J. (2004). Functional connectivity patterns of human magnetoencephalographic recordings: a 'small-world' network? *Neurosci. Lett.* 355, 25–28. doi: 10.1016/j.neulet.2003.10.063

Stam, C. J., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* 28, 1178–1193. doi: 10.1002/hbm.20346

Stam, C. J., and Van Straaten, E. C. (2012). The organization of physiological brain networks. *Clin. Neurophysiol.* 123, 1067–1087. doi: 10.1016/j.clinph.2012.01.011

Stephan, K. E., and Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *Wiley. Interdiscip. Rev. Cogn. Sci.* 1, 446–459. doi: 10.1002/wcs.58

Tahmasian, M., Bettray, L. M., Van Eimeren, T., Drzezga, A., Timmermann, L., Eickhoff, C. R., et al. (2015). A systematic review on the applications of resting-state fMRI in Parkinson's disease: does dopamine replacement therapy play a role? *Cortex* 73, 80–105. doi: 10.1016/j.cortex.2015.08.005

Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., and Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. *Nature* 575, 137–146. doi: 10.1038/s41586-019-1657-6

Tremblay, P., and Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* 162, 60–71. doi: 10.1016/j.bandl.2016.08.004

Uhlhaas, P. J., Roux, F., Rodriguez, E., Rotarska-Jagiela, A., and Singer, W. (2010). Neural synchrony and the development of cortical networks. *Trends Cogn.* 14, 72–80. doi: 10.1016/j.tics.2009.12.002

Van Diessen, E., Numan, T., Van Dellen, E., Van Der Kooi, A. W., Boersma, M., Hofman, D., et al. (2015). Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. *Clin. Neurophysiol.* 126, 1468–1481. doi: 10.1016/j.clinph.2014.11.018

Van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., et al. (2014). Functional brain connectivity from EEG in epilepsy: seizure prediction and epileptogenic focus localization. *Prog Neurobiol* 121, 19–35. doi: 10.1016/j.pneurobio.2014.06.004

Vasil'yeva, N. O., and Shmalei, S. V. (2013). Coherent relations in ongoing encephalograms of preschool boys with neurosis-like stammering. *Neurophysiology* 45, 468–476. doi: 10.1007/s11062-013-9396-z

Vértes, P. E., and Bullmore, E. T. (2015). Annual research review: Growth connectomics–the organization and reorganization of brain networks during normal and abnormal development. *J. Child Psychol. Psychiatry* 56, 299–320. doi: 10.1111/jcpp.12365

Weiss, S., and Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Front. Psychol.* 3:201. doi: 10.3389/fpsyg.2012.00201

Weiss-Croft, L. J., and Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22years of functional MRI. *Neuroimage* 123, 269–281. doi: 10.1016/j.neuroimage.2015.07.046

Whedon, M., Perry, N. B., Calkins, S. D., and Bell, M. A. (2016). Changes in frontal EEG coherence across infancy predict cognitive abilities at age 3: the mediating role of attentional control. *Dev. Psychol.* 52, 1341–1352. doi: 10.1037/dev0000149

Williams, I. A., Tarullo, A. R., Grieve, P. G., Wilpers, A., Vignola, E. F., Myers, M. M., et al. (2012). Fetal cerebrovascular resistance and neonatal EEG predict 18-month neurodevelopmental outcome in infants with congenital heart disease. *Ultrasound Obstet. Gynecol.* 40, 304–309. doi: 10.1002/uog.11144

Yang, C. C., Yang, C. C., and Chaou, W. T. (2005). Functional correlations of spatial quantitative EEG and intelligences in a nonalphabetical language group. *Appl. Neuropsychol.* 12, 151–157. doi: 10.1207/S15324826AN1203_5

Youssofzadeh, V., Williamson, B. J., and Kadis, D. S. (2017). Mapping critical language sites in children performing verb generation: whole-brain connectivity and graph theoretical analysis in MEG. *Front. Hum. Neurosci.* 11:173 doi: 10.3389/fnhum.2017.00173

Zare, M., Rezvani, Z., and Benasich, A. A. (2016). Automatic classification of 6-month-old infants at familial risk for language-based learning disorder using a support vector machine. *Clin. Neurophysiol.* 127, 2695–2703. doi: 10.1016/j.clinph.2016.03.025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.