Brain and Cognition 151 (2021) 105730

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

The electrophysiology of audiovisual processing in visual narratives in adolescents with autism spectrum disorder

Mirella Manfredi [a,c,*], Neil Cohn [b], Beatriz Ribeiro [a], Pamella Sanchez Pinho [a], Elisabete Fernandes Rodrigues Pereira [a], Paulo Sergio Boggio [a]

[a] Social and Cognitive Neuroscience Laboratory, Center for Biological Science and Health, Mackenzie Presbyterian University, São Paulo, Brazil
[b] Department of Communication and Cognition, Tilburg University, Tilburg, Netherlands
[c] Department of Psychology, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Autism
Multimodality
N400
Audiovisual
Visual narrative

ABSTRACT

We investigated the semantic processing of the multimodal audiovisual combination of visual narratives with auditory descriptive words and auditory sounds in individuals with ASD. To this aim, we recorded ERPs to critical auditory words and sounds associated with events in visual narrative that were either semantically congruent or incongruent with the climactic visual event. A similar N400 effect was found both in adolescents with ASD and neurotypical adolescents (ages 9–16) when accessing different types of auditory information (i.e. words and sounds) into a visual narrative. This result might suggest that verbal information processing in ASD adolescents could be facilitated by direct association with meaningful visual information.

In addition, we observed differences in scalp distribution of later brain responses between ASD and neurotypical adolescents. This finding might suggest ASD adolescents differ from neurotypical adolescents during the processing of the multimodal combination of visual narratives with auditory information at later stages of the process. In conclusion, the semantic processing of verbal information, typically impaired in individuals with ASD, can be facilitated when embedded into a meaningful visual information.

1. Introduction

Individuals with Autism Spectrum Disorder (ASD) have documented struggles with both auditory processing (for a review, see O'Connor, 2012) and language processing (Coderre, 2020). Contrastingly, it is often assumed that picture processing remains intact, despite observations that processing visual communication also faces deficits, particularly in sequential contexts like sequences of images in comics or picture stories (Coderre, 2020). Here we seek to explore this relationship between verbal and visual processing by examining their *interaction*. Specifically, we asked whether the semantic processing of the multimodal audiovisual combination of visual narratives (comic strips) with auditory descriptive words would differ from a combination with auditory sounds in individuals with ASD.

To investigate semantic processing, previous studies on both neurotypical and clinical populations have measured event-related brain potentials (ERP), particularly focusing on the N400, an ERP component thought to index semantic processing. This ERP component peaks roughly 400 ms after the onset of a stimulus and is thought to index the spreading activation that occurs in the access of semantic information by a stimulus in relation to its preceding context (Kutas & Federmeier, 2011). The amplitude of this component is greater to critical words that are semantically unrelated (vs. related) to previous single words in semantic priming paradigms (Bentin et al., 1985; Kutas & Hillyard, 1980). Thus, a greater N400 response appears to words that are incongruent (vs. congruent) with their preceding sentence contexts (Kutas & Hillyard, 1980, 1984) or global discourse contexts (Camblin et al., 2007).

The N400 also appears to semantic processing outside of language, but both the morphology and the latency of the N400 response appear to be sensitive to differences in sensory modality. For instance, previous studies suggested that sounds and words evoked an N400 effect, but the distribution and the latency of the N400 effect to words differed from that of sounds (Van Petten & Rheinfelder, 1995). Specifically, they observed a right dominant hemispheric laterality for words and left dominant laterality for environmental sounds.

The N400 also occurs in response to individual images (Van Berkum

* Corresponding author.
E-mail address: mirella.manfredi@uzh.ch (M. Manfredi).

https://doi.org/10.1016/j.bandc.2021.105730
Received 3 November 2020; Received in revised form 15 February 2021; Accepted 3 April 2021
0278-2626/Published by Elsevier Inc.

et al., 2003; Barrett & Rugg, 1990), visual sequences (Sitnikova et al., 2008) or visual narratives like comics (Cohn, 2012). However, visual information typically elicits a more widespread frontal distribution (Barrett & Rugg, 1990; Ganis et al., 1996) compared to written words (Kutas and Hillyard, 1984).

The N400 even appears across interactions between modalities: some previous studies have examined semantic processing by replacing words in sentences (Ganis et al., 1996), or words replacing panels in comics (Manfredi et al., 2017). These studies revealed that although the critical units switched the modality of their context, the distribution of the N400 effects was more similar to those that appeared when keeping the modality constant: the N400s to pictures in sentences had a frontal distribution like image-N400s, while words in visual sequences had distribution like word-N400s. These results imply that, despite sharing similar mechanisms of processing (N400), the cortical areas of origin may have some sensitivity to different types of stimuli.

Other studies have investigated semantic processing by combining stimuli presented simultaneously in different sensory modalities (i.e., vision and sound). For example, previous work observed N400 effects when speech and/or natural sounds were combined with semantically inconsistent pictures or video frames (Cummings et al., 2008; Liu et al., 2011; Plante et al., 2000; Puce et al., 2007). Similar results arose when gestures were combined with (in)congruent verbal information (Cornejo et al., 2009; Coulson & Wu, 2005a, 2005b; Proverbio et al., 2014; Wu & Coulson, 2007). The congruity of gesture-music pairings also affects N400 amplitudes, in musicians (Proverbio et al., 2014). Thus, congruity can occur not only with a stimulus and its sequential context, but also its simultaneous interactions with other modalities.

In another recent ERP study (Manfredi et al., 2018), we examined the multimodal interactions between visual narratives and auditory words and sounds, which were either semantically related or unrelated to the climactic visual event. For example, a depiction of Snoopy spitting up a ball could accompany the congruent sound of spitting or the word "spitting" or the incongruent sound of something hitting something else or the word "hitting", as in Fig. 1. We observed that both incongruent sounds and words evoked an N400 effect, but the distribution and latency of the N400 effect differed between modalities. However, a sustained late frontal negativity did not differ between types of stimuli, supporting the idea that semantic processing balances a distributed cortical network accessible from multiple modalities, yet also involves more general mechanisms insensitive to specific modalities.

Deficits in semantic processing have been observed in ASD. In particular, it has long been claimed that individuals with autism face deficits in comprehending verbal materials. Previous ERP studies have found attenuated N400 effect in the ASD population, reflecting deficits in neural processing of verbal semantic information (Braeutigam et al., 2008; Dunn et al., 1999; Dunn & Bates, 2005; Kujala et al., 2013; O'Connor, 2012; Pijnacker et al., 2010); Lepistö et al., 2005; Fishman et al., 2011). For example, Pijnacker et al. (2010) studied the processing of sentence context and reasoning context in adults with high-functioning autism (HFA), Asperger syndrome, and in a matched control group. Their results revealed that adults with HFA required more elaborate processing for sentence interpretation. Braeutigam and colleagues (2008) performed a MEG study in which found a weaker N400 effect in response to incongruous words for individuals with autism compared with a control group.

Despite these issues with linguistic processing in ASD, according - to widespread thinking, the comprehension of visual materials is presumed to remain intact, a belief described as the "Visual Ease Assumption" (for review, see Coderre, 2020). Because of this assumption, researchers frequently use experimental tasks with visual narratives like comics because they are thought to be easier for children with ASD to process than verbal stimuli (Coderre, 2020).

A growing literature has shown that visual narratives are governed by structural constraints analogous to those found in written sentences (Cohn, 2020; Cohn et al., 2012). Like in sentences, the processing of meaning in visual narratives appears to involve a cyclic relationship of semantic access and subsequent integration or updating, as indexed by the N400 and P600 ERPs, respectively (Cohn, 2019; Manfredi et al., 2017, 2018). In the first study examining ERPs between language and visual narratives related to individuals with ASD, Coderre and colleagues (Coderre et al., 2018) compared semantic comprehension of short verbal narratives where the final word was either congruous or incongruous, and they also compared short 6-panel visual narratives (i.e., comic strips) with congruous or incongruous final images. Attenuated N400 effects were evoked by linguistic narratives for the ASD population

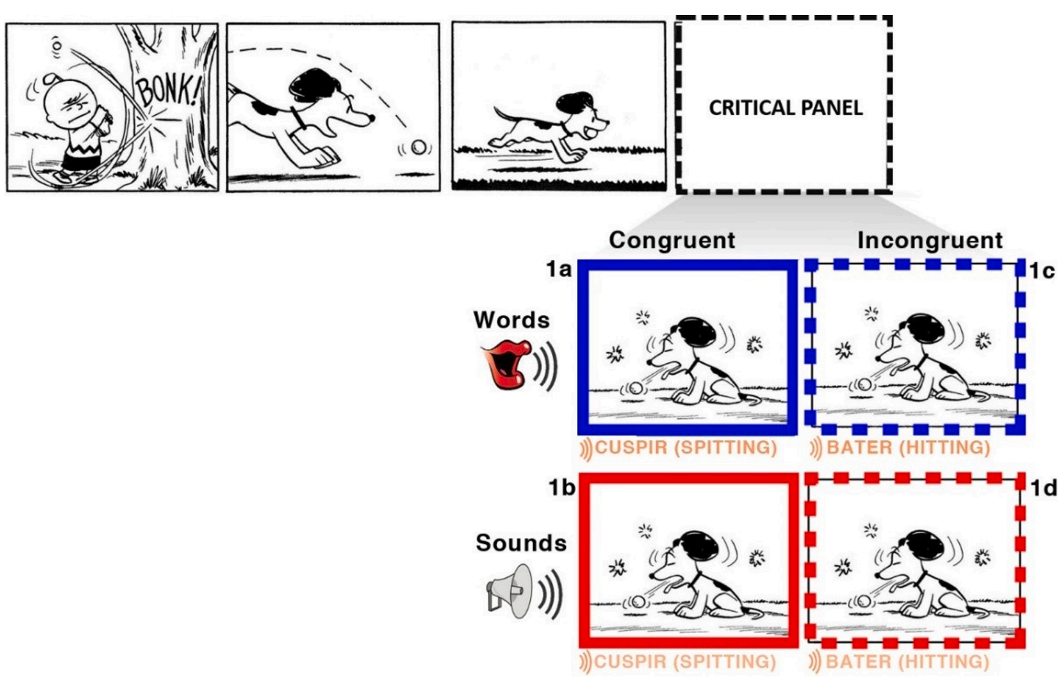

Fig. 1. Example of visual sequences used as experimental stimuli. We manipulated these base sequences by adding a congruent word, an incongruent word, a congruent sound and incongruent sound.

compared to the controls, but so were the N400 effects to the visual narratives. Also, unlike prior studies of typically developing adults (Cohn et al., 2012), the N400s to coherent narratives for individuals with ASD showed no attenuation across the ordinal position of these sequences. These findings implied that difficulties shown by individuals with ASD may not be due to a modality-specific verbal impairment.

Similar to Coderre et al. (2018), our group (Manfredi et al., 2020) investigated the semantic processing of auditory sentences and visual narratives in adolescents with ASD and typically-developing (TD) adolescents (ages 9–16). We observed a focal central N400 effect to incongruent words as compared to congruent ones, which was slightly attenuated for ASD than TD adolescents. A subsequent negativity sustained to incongruous than congruous words, but only for the TD adolescents, possibly reflecting a cost of further processing the inconsistent auditory information that was absent for adolescents with ASD. Incongruent critical panels in visual narratives evoked a greater fronto-central N400 than congruent panels in both groups. In addition, incongruent panels evoked a centro-parietal late positivity, but only for the TD adolescents, possibly reflecting that the adolescents with ASD may not have been as sensitive to integrating the discontinuities of the incoming visual information. These findings suggest that adolescents with ASD face processing deficits in both verbal and visual materials when integrating meaning across sequential units, though such impairments may arise in different parts of the interpretive process, depending on the modality.

Few studies have focused on the semantic processing of the audiovisual stimuli in ASD, though some previous works have compared verbal and non-verbal information (McCleery et al., 2010; Ribeiro et al., 2013). McCleery et al. (2010) recorded EEG while children attended to semantically matching and mismatching picture–word and picture–environmental sound pairs. Typically developing children exhibited evidence of semantic incongruency detection (N400) in both the word and environmental sound conditions. In contrast, children with ASD showed an N400 effect to the environmental sounds but not the words. These results indicate an impairment in the automatic activation of semantic representations in children with ASD and suggest that this deficit could be more severe in the verbal than the nonverbal domain.

Other studies have compared cognitive mechanisms evoked by linguistic versus nonlinguistic tasks. Mongillo et al. (2008) compared ASD children and neurotypical children on six perceptual tasks. Children with ASD scored significantly lower than children without ASD on audiovisual tasks involving human faces and voices but scored similarly to children without ASD on audiovisual tasks involving nonhuman stimuli (bouncing balls). These results supported previous findings that children with autism are less influenced by visual speech information (Massaro & Bosseler, 2003) and they may use visual information for speech differently. This evidence suggests that the audiovisual integration of nonverbal stimuli in ASD might be comparable to that observed in matched controls. On the other hand, these findings might also suggest a deficit specific to the verbal domain during the audiovisual perception in ASD (McCleery et al., 2010).

To our knowledge, no previous studies have investigated the multimodal interactions between visual narratives and auditory words and sounds. In the present study, we investigated the brain responses of neurotypical and adolescents with ASD when presenting visual narrative together with different types of auditory information, such as sounds and words (Manfredi et al., 2018). This study investigated whether the semantic processing of the multimodal audiovisual combination of visual narratives (comic strips) with auditory descriptive words would differ from a combination with auditory sounds in individuals with ASD. Because of the sequential nature of the images, this design allows us to naturalistically examine whether the image sequence semantically primes the ASD children's response to words and sounds in a naturalistic way.

It is important to note that in the current study, the participants were the same as in this previous one on unimodal visual narratives (Manfredi et al., 2020). Therefore, the results of this investigation would complement those observed previously by focusing on the semantic processing of verbal information embedded in a multimodal interaction in a visual narrative, in contrast to the differences observed in unimodal verbal/visual narratives.

We recorded ERPs to auditory words and sounds associated with events in visual narratives—i.e., seeing images of someone spitting while hearing either a word or a sound, as in Fig. 1 (see Manfredi et al., 2017). Auditory words and sounds were either semantically congruent or incongruent with the climactic visual event. Based on the previous findings (Coderre et al., 2018), we expected that if modality does not interact with the deficits in ASD, then both words and sounds should be equally affected.

2. Methods

2.1. Participants

The experimental group included twenty-four ASD school-aged children and adolescents (mean age = 11.4, SE = 1.9) and sixteen age-matched typically developing children and adolescents (mean age = 12.6, SE = 2.1). All participants were right-handed male, Portuguese speakers with normal or corrected-to-normal vision. All the participants' parents provided written informed consent. The study was approved by the Institutional Ethics Committees of the Mackenzie Presbyterian University and by the National Ethics Committee of Brazil. The participants were the same as in our previous work (Manfredi et al., 2020).

All of the children with autism met the DSM-IV criteria for Autism Spectrum Disorder (American Psychiatric Association, 1994) and were diagnosed through structured assessment by a developmental psychiatrist, and further confirmed using the Autism Diagnostic Interview-Revised (Lord et al., 1994). Autistic children were recruited from three health centers located in Sao Paulo: *Apace, APRAESPI, Santa Casa*. Controls were recruited from their immediate community members to minimize differences in socioeconomic status.

An abbreviated version (2 subsets; Cube and Vocabulary) of the Brazilian WISC-III (WISC-III; Wechsler, 1992) was also administered to calculate an estimated IQ (de Mello et al., 2011). No significant differences were found between the estimated mental age of the autistic children (M = 100, SE = 11.39) and the control group (M = 107, SE = 18.88). Exclusionary criteria for all the groups included a nonverbal IQ below 80, as assessed by the *WISC-III* (Wechsler, 1999) and a history of significant psychiatric or neurological comorbidities (see Table 1).

Twelve ASD participants and three controls were excluded from the ERP statistical analyses because of EEG artifacts. This left 12 final participants in the ASD group, with 13 final participants in the neurotypical control group.

2.2. Stimuli

We designed 100 novel 4 panel long visual narrative sequences using black and white panels from the *Complete Peanuts* volumes 1 through 6 (1950–1962) by Charles Schulz (Fantagraphics Books, 2004–2006). We used the same stimulus set we have used in the previous study with neurotypical adults (Manfredi et al., 2018). To eliminate the influence of written language, we used panels either without text or with text deleted. All panels were adjusted to a single uniform size. All sequences had a coherent narrative structure, as defined by the theory of Visual Narrative Grammar (Cohn, 2013), and confirmed by behavioral ratings (Cohn et al., 2012).

We combined these base sequences with a sound or an auditory word time-locked to the image of the narrative climax of each strip, following the same method as in our previous research with neurotypical adults (Manfredi et al., 2018). Congruent sequences matched climatic pictures with auditory Portuguese words that described familiar actions or with sounds that corresponded to the same actions. For example, as in Fig. 1a,

Table 1
Participant characteristics for the TD and ASD groups. Means and ranges are reported for each measure. The 'group difference' column shows the results of independent-samples t-tests on each measure. We reported participant characteristics for both the total sample and the analyzed sample.

		TOTAL SAMPLE	
	CONTROL GROUP	ASD GROUP	GROUP DIFFERENCE
AGE	12.6 (9–16)	11.4 (9–15)	p = n.s.
SEX	males	males	
IQ total	107 (88–132)	100 (88–120)	p = n.s.
IQ verbal	12	7.8	p < 0.05
IQ performance	13.5	12.4	p = n.s.
Type of school	Public	Public	
		ANALYZED SAMPLE	
	CONTROL GROUP	*ASD GROUP*	*GROUP DIFFERENCE*
AGE	12.7 (10–16)	11.3 (9–15)	p = n.s.
SEX	males	males	
IQ total	108 (88–132)	101 (88–120)	p = n.s.
IQ verbal	12	8	p < 0.05
IQ performance	12.7	12.2	p = n.s.
Type of school	Public	Public	

an image of a dog spitting might appear concurrent to the auditory sound of the word *cuspir* ("spitting"). We presented 27 auditory Portuguese words and 23 sounds. The latter included environmental (n = 11), human (n = 9) and animal sounds (n = 3). We then created four sequence types by modulating the type of stimulus (words vs. sounds) and their congruence with the visual images (congruent vs. incongruent). As in Fig. 1, *congruent word* panels (1a) contained an auditory word coherent with the contents of the image, *congruent sound* panels (1b) used a sound coherent with the image, *incongruent word* panels (1c) contained an auditory word incoherent with the image, and *incongruent sound* panels (1d) used a sound incoherent with the image. The critical panels appeared in the second to the fourth-panel positions, with equal numbers at each position. Some words and sounds were repeated across the sequences but with different comic strips. The average values (number of repetitions) were not significantly different across the conditions ($t(47) = 1.10$, $p > 0.05$) (Words = 4.43, SD = 1.40; Sounds = 3.92, SD = 1.78).

A female native Portuguese speaker produced the word stimuli (mean duration = 996 ms, SD = 72.01 ms), which were recorded in a single session in a sound-attenuating booth. Environmental sound stimuli (mean duration = 985 ms, SD = 42.31 ms) were obtained from several online sources. Word stimuli and environmental sounds were standardized for sound quality (44.1 kHz, 16 bit, stereo). A *t*-test revealed no differences (p = 0.5) between the duration of words (996 ms, SD = 72.01 ms) and environmental sounds (985 ms, SD = 42.31 ms).

Pre-assessment of stimuli were made by a group of 8 judges of similar age and educational levels as the control participants, but who did not participate in the EEG study. Congruent sequences rated as incoherent by at least 80–99% of judges were discarded, as were incongruent sequences evaluated as coherent. Our final stimulus set included 100 experimental sequences (~25 per condition). A total of four lists (each consisting of 100 strips in random order) were created, with the four conditions counterbalanced using a Latin Square Design such that participants viewed each sequence only once in a list.

2.3. Procedure

Participants sat in front of a monitor in a sound-proof, electrically-shielded recording chamber. During the visual presentation of stimuli, each panel stayed on screen for 1350 ms, separated by an ISI of 500 ms. When the strip concluded, the question "was it understandable?" (in Portuguese) appeared on the screen, and participants responded by pressing one of the two hand-held buttons. Response hand was counterbalanced across participants and lists. The experiment had four blocks separated by breaks, and experimental trials were preceded by a short practice to familiarize participants with the procedures.

2.4. Electroencephalographic recording parameters

The electroencephalogram (EEG) was recorded from 128 electrodes at a sampling rate of 250 Hz (bandpass 0.01–100 Hz). The EEG was recorded and analyzed using the Net station software (*Geodesic EEG Net Station, EGI*, Eugene, OR). The impedance of all electrodes was kept below 50 kΩ over the experiment. All recordings were referenced to Cz electrode during data acquisition. This allowed us to analyze the mastoid-temporal lobe activity in addition to all other important sites for linguistic processing. EEG epochs were synchronized with the onset of stimuli presentation (Fig. 2).

2.5. Statistical analysis of ERPs

Trials contaminated by blinks, muscle tension (EMG), channel drift, and/or amplifier blocking were discarded before averaging. Approximately 9% of critical panel epochs were rejected due to such artifacts, with losses distributed approximately evenly across the four conditions. Each participant's EEG was time-locked to the onset of critical panels and their accompanying auditory stimuli. ERPs were computed for epochs extending from 100 ms before stimulus onset to 1500 ms after stimulus onset.

These responses were measured at 72 electrode sites (8 in each region) in fronto-left (23, 24, 26, 32, 33, 27, 34, 28), fronto-central (18, 16, 10, 22, 9, 15, 21, 14), fronto-right (3, 124, 2, 123, 122, 1, 117, 116), centro-left (39, 40, 45, 46, 41, 47, 50, 51), central (7, 106, 31, 80, 55, 30, 105, 79), centro-right (104, 103, 102, 98, 101, 97, 115, 109), parieto-left (58, 59, 60, 70,66, 65, 69, 64), parietal (61, 62, 78, 67, 72, 77, 71, 76), parieto-right (96, 85, 91, 83, 84, 90, 95, 89) electrode sites.

The mean amplitude of N400 (350–550 ms) and *Late Positivity (550–750 ms)* were analyzed using repeated measures ANOVAs with Group (2 levels: ASD children, Controls), Modality (2 levels: words, sounds), Congruity (2 levels: Congruent, Incongruent), AP Distributions (3 levels: Frontal, Central, Parietal) and Laterality (3 levels: Left, Medial, Right).

Multiple comparisons of means were performed with post-hoc Fisher's tests.

3. Results

3.1. Behavioral results

Overall, a 2 × 2 ANOVA computed on ratings on the two groups revealed a significant main effect of Congruency ($F(1, 18) = 7.55$, $p < 0.01$), arising because sounds/words which were congruent with the visuals were rated as more coherent (71%, SE = 15.22) that the incongruent pairings (35%, SE = 18.06). There was no main effect of Modality or Group (p = n.s.).

3.2. Electrophysiological results

3.2.1. N400 (350–550 ms)

In the N400 time window, the omnibus ANOVA revealed a significant interaction of Modality × AP Distribution × Group [($F(2, 46) = 3.05$, $p < 0.05$)] suggested a greater negativity to words (7.07 μV, SE = 1.15) compared to sounds (8.84 μV, SE = 1.20) only in the parietal areas in the control group ($p = 0.05$). No differences were found between sounds (7.09 μV, SE = 1.25) and words (7.26 μV, SE = 1.20) in the ASD group (Fig. 3).

Moreover, the significant interaction of Modality × Congruency ×

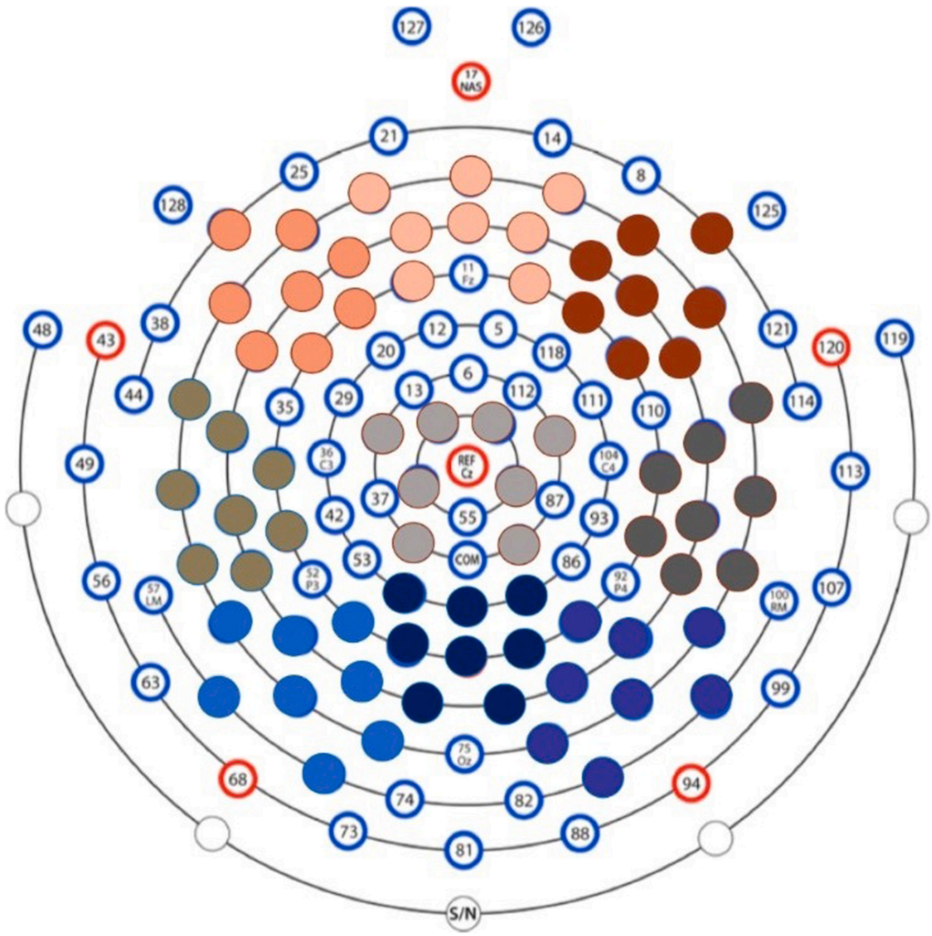

Fig. 2. Schematic showing the 128 channel array of scalp electrodes from which the EEG was recorded.

AP Distribution × Electrode [(F (14, 322) = 2.56, $p < 0.05$)] revealed a greater N400 response to incongruent words (1.27 μV, SE = 0.56) compared to congruent ones (2.98 μV, SE = 0.68) only at four electrodes located on the central regions (all $p < 0.05$) and to incongruent sounds compared to congruent ones on four electrodes on the frontal sites (incongruent sounds: −6.11 μV, SE = 0.77; congruent sounds: −4.63 μV, SE = 0.70), two electrodes on the central area (incongruent sounds: −6.01 μV, SE = 0.65; congruent sounds: 1.63 μV, SE = 0.58) and six electrodes on the parietal areas (incongruent sounds: 6.94 μV, SE = 1.20; congruent sounds: 8.96 μV, SE = 0.96; all $p < 0.05$) (Figs. 3 and 4).

3.2.2. *Later effects (550–750 ms)*

Additional analysis in the later time window revealed a significant interaction of Modality × AP Distribution × Laterality [(F (4, 92) = 2.74, $p < 0.05$)], suggesting a greater sustained negativity to words (-4.15 μV, SE = 0.56) compared to sounds (-2.56 μV, SE = 0.51) in the fronto-right areas and a greater LP to sounds (3.07 μV, SE = 0.65) compared to words (2.01 μV, SE = 0.76) in the centro-medial, centro-right and right-parietal sites (Figs. 3 and 4).

In addition, the significant interaction of Congruency, AP Distribution, Laterality and Group [(F (4, 92) = 2.61, $p < 0.05$)], indicated a greater LP response to incongruent (6.37 μV, SE = 1.25) than congruent stimuli (3.65 μV, SE = 1.14) in the centro-left, parieto-central and parieto-right sites in the ASD group, and a greater sustained negativity to incongruent (3.04 μV, SE = 1.23) than congruent stimuli (4.42 μV, SE = 0.98) only in the centro-medial sites and centro-parietal sites in the control group (Fig. 5).

4. Discussion

In this study, we investigated whether the semantic processing of the multimodal combination of visual narratives with auditory descriptive words differed from that with auditory sounds in individuals with ASD. We recorded ERPs to critical auditory words and sounds associated with events in a visual narrative that were either semantically congruent or incongruent with the climactic visual event. Our findings did not reveal substantial differences in ASD individuals' brain responses (i.e., N400 effect) when accessing and integrating different kinds of auditory information into a visual narrative. In particular, our results showed no differences in verbal-visual semantic processing in the two groups of children, suggesting that verbal information processing in ASD adolescents could be facilitated by direct association with meaningful visual information (as we hypothesized). However, varying scalp distributions of later brain responses between ASD and control children might suggest differences during the processing of the multimodal combination of visual narratives with auditory information (words and sounds) at different stages of the interpretive process.

Our study revealed a greater N400 response to both the incongruent words compared to congruent ones and to the incongruent sounds compared to congruent ones, without any difference between the two groups of children. This larger N400s to incongruent auditory stimulus with visual information is consistent with previous studies of larger N400s to unexpected or semantically incongruent information (Kutas & Federmeier, 2011). Previous studies found N400 effects to anomalies in unimodal contexts to words in sentences (e.g., Bentin, et al., 1985; e.g., Kutas & Hillyard, 1980, 1984; Camblin, et al., 2007), words embedded in discourse (Van Berkum et al., 1999; 2003), multimodal relationships

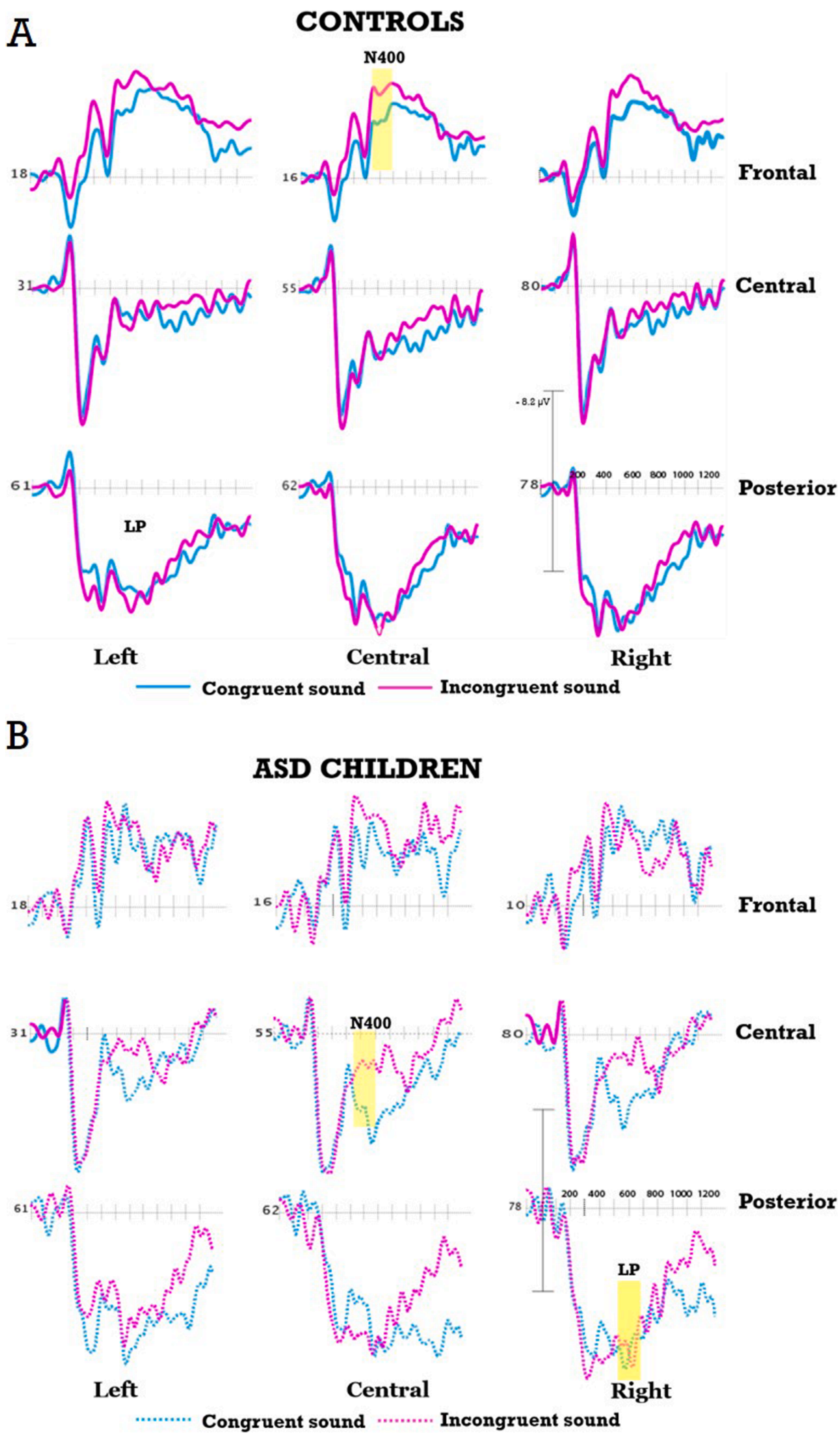

Fig. 3. **(A)** Grand-average ERP waveforms recorded at central sites in response to Congruent words (blue) and Incongruent words (red) in the control group. **(B)** Grand-average ERP waveforms recorded at central sites in response to Congruent words (blue) and Incongruent words (red) in the ASD group. Time windows of interest for the ERP components are highlighted in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between speech and gesture (Cornejo et al., 2009; Habets et al., 2011; Özyürek et al., 2007; Proverbio et al., 2014; Wu and Coulson, 2005; 2007a; 2007b) and between speech or natural sounds with pictures or videos (Cummings et al., 2008; Liu et al., 2011; Plante et al., 2000; Puce et al., 2007). This N400 effect to words had a central distribution of the incongruity effect, similar to that observed in prior work with adults (Manfredi et al., 2018). These findings indicate that the N400 effect might have the same neural generator source in both adolescents and adults.

The greater negativities to both groups differ from previous ERP studies which have revealed deficits in neural processing of unimodal verbal semantic information for individuals with ASD (Dunn and Bates, 2005; Dunn et al., 1999; Pijnacker etal., 2010; Kujala etal., 2013; Braeutigam etal., 2008; Lepistö etal., 2005; Fishman etal., 2011; see O'Connor 2012). For example, in a MEG study, Braeutigam et al. (2008) found an attenuated N400-type effect in response to incongruous words

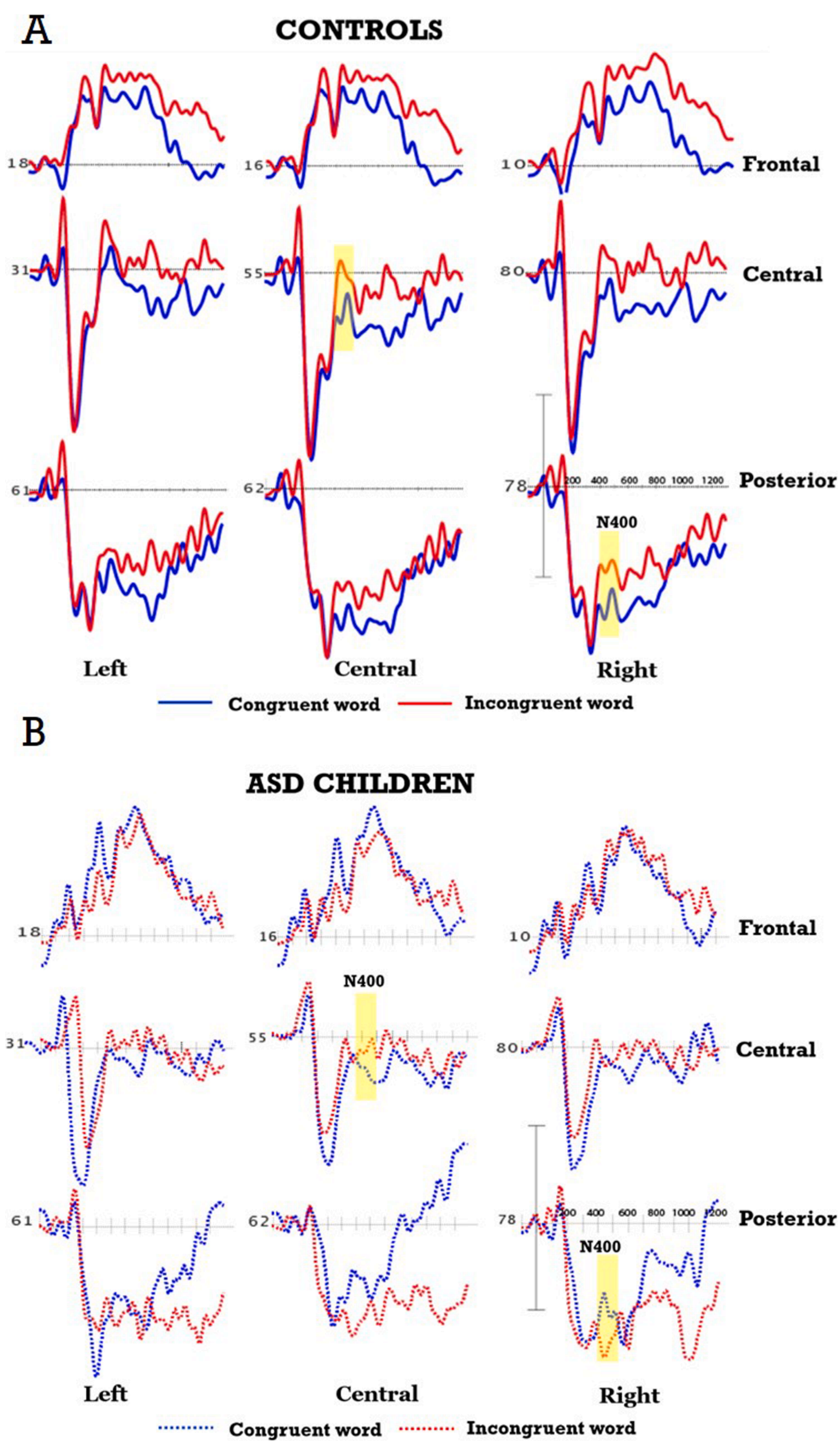

Fig. 4. **(A)** Grand-average ERP waveforms recorded at central sites in response to Congruent sounds (light blue) and Incongruent sounds (lilac) in the control group. **(B)** Grand-average ERP waveforms recorded at central sites in response to Congruent sounds (light blue) and Incongruent sounds (lilac) in the ASD group. Time windows of interest for the ERP components are highlighted in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in individuals with autism.

In addition, a previous ERP study by our group (Manfredi et al., 2020) observed attenuated N400s to incongruent words in ASD adolescents, compared to the control group, whereas incongruent sequence-final images in visual narratives evoked an increased N400 effect in both groups. In the current study, we observed no differences in semantic processing of verbal information in the two groups of children. It is worth noting that the participants in the current study were the same as in this previous one on unimodal visual narratives (Manfredi et al., 2020). Thus, the present results corroborate and complement those observed previously, suggesting that the semantic processing of verbal information is facilitated for these ASD adolescents when embedded in a

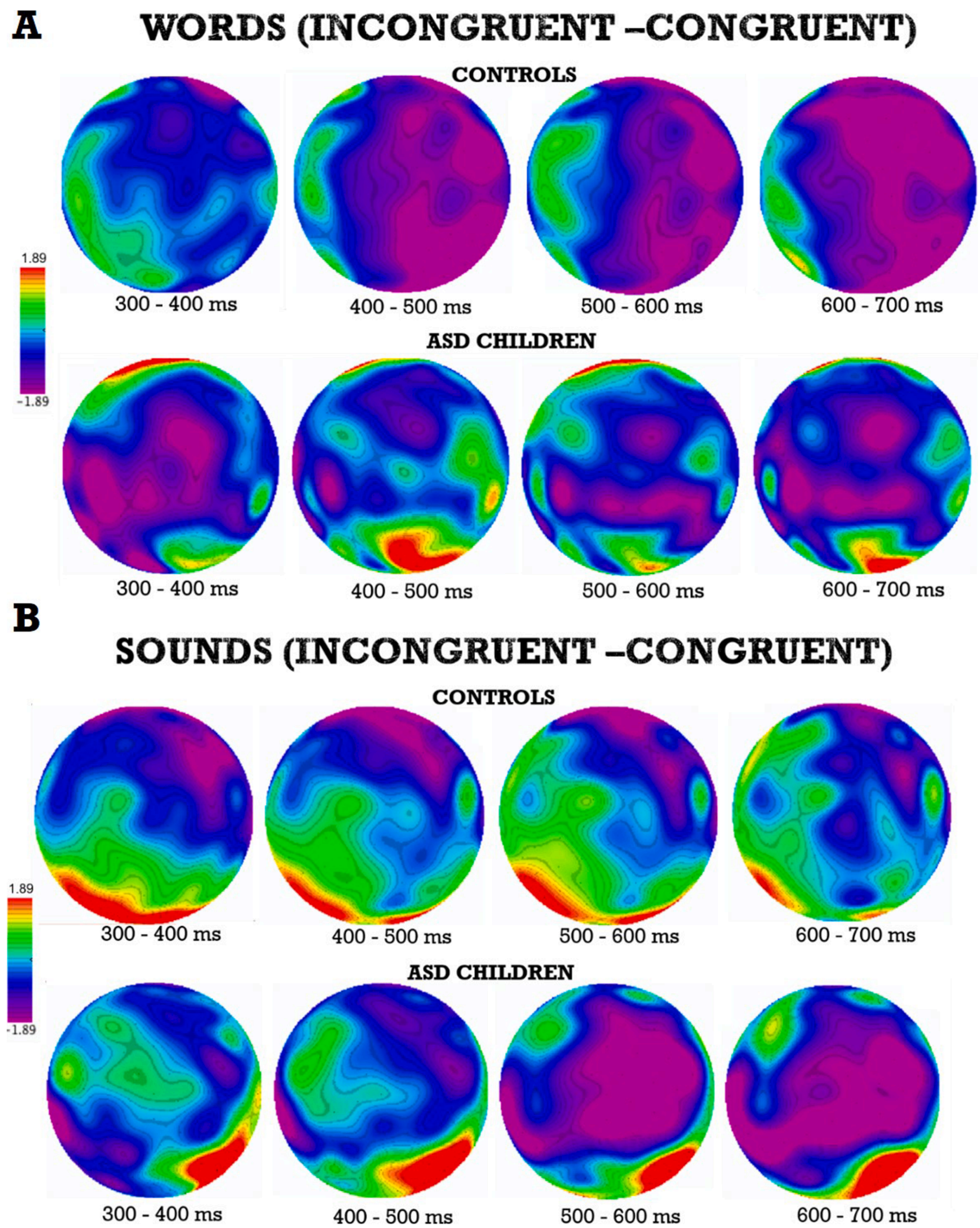

Fig. 5. **(A)** Voltage of scalp distribution of the N400 in response to the difference between congruent and incongruent words in control groups (first row) and ASD (second row). **(B)** Voltage of scalp distribution of the N400 in response to the difference between congruent and incongruent sounds in control groups (first row) and ASD (second row).

multimodal interaction in a visual narrative, in contrast to the differences observed in unimodal verbal/visual narratives.

An N400 effect also appeared to incongruent sounds interacting with the visual narratives in both the groups of children. Our results are similar to those observed in a previous EEG study by McCleery et al. (2010) in which children attended to semantically matching and mismatching picture–word and picture–environmental sound pairs. While they observed differences in words processing, no differences emerged in environmental sounds processing between the two groups of children. Thus, our study further suggested that semantic deficits in adolescents with ASD does not appear to affect the auditory domain in multimodal interactions. As we expected, our findings revealed that the multimodal combination of visual narrative with sounds does not appear to be impaired in ASD children.

Finally, the incongruity effect to sounds in children differs from the group of adults in a prior study (Manfredi et al., 2018) that showed a fronto-central distribution to sounds, while the group of children in the current study showed a widespread effect across the scalp. It is possible that in children sounds don't elicit a selective activation as words. Maybe because children don't associate the environmental sound to a specific source yet and so they may exert greater effort and thus may recruit more widespread cortical activation than adults when integrating environmental sounds into the context of a visual narrative.

Following the N400, we observed a greater LP response to sounds compared to words in the centro-parietal sites. In addition, the analysis revealed a greater LP response to incongruent than congruent stimuli in the centro-left, parieto-central and parieto-right sites for the adolescents with ASD, and a greater LP response to incongruent than congruent stimuli only in the centro-medial sites and centro-parietal sites in the typically developing adolescents. The function of the posterior LP has been associated with updating or reanalysis processes (Van Petten & Luka, 2012), especially when the incoming word disconfirms predictions based on a preceding context (Quante et al., 2018; Kuperberg, 2013).

In our previous study (Manfredi et al., 2020) these typically developing adolescents showed a greater P600 with a centro-parietal distribution during the presentation of incongruent panels in a visual narrative, while these adolescents with ASD showed an attenuated LP. This late response suggested that only the neurotypical children recognized the discontinuity of the incoming visual information relative to its prior context. In contrast, here, both groups, which were the same participants as that prior study, showed a larger LP response for incongruous than congruent stimuli. This finding might suggest that the semantic processing of visual narrative benefits from the multimodal interaction created by a direct association of the visual content with congruous auditory information. Thus, in unimodal context these adolescents with ASD evoked less updating of information across the panels of the visual narrative, while in a multimodal context, updating remained necessary for the interaction of visual and auditory information.

Nevertheless, in addition to this late positivity, a greater sustained fronto-central negativity appeared in the 550–750 ms time window to words compared to sounds in the frontal areas, and to incongruous than congruous stimuli. If this sustained frontal negativity is consistent with prior findings, it may reflect a general cost of further processing the inconsistent interaction between auditory information and visual narrative (Manfredi et al., 2018). Here, the sustained negativity could be an index of the sustained interpretative processing within a mental model, while the late positivity could reflect the revision or update of the model (Baggio et al., 2008).

5. Conclusion

Overall, in this study, we found that the semantic processing of verbal information in adolescents with autism was facilitated by direct association with meaningful visual information. This raises questions about the degree to which semantic processing benefits from multimodality in ASD in contrast to unimodal processing. An additional question is at which stage of development this interaction may become beneficial. Is this facilitation similar in children with autism and in healthy children? It would thus be interesting to study the brain responses in healthy children and in children with autism at different ages. Future studies might analyze how verbal information interacts with visual events in the context of a visual narrative sequence at different stages of development and verify whether, and if so when, this interaction differs between typically developed children and clinical populations, such as autism.

6. Authors statement

MM and PSB ideated and designed the research. MM and NC constructed the materials and collected the data. BR, PSM and EF technically supported data collection. MM and NC analyzed the data. MM, NC and PSB interpreted the data and wrote the first draft of the manuscript. All authors contributed to this article and approved the final version of the manuscript.

7. Disclosure statement

No potential conflict of interest was reported by the authors

Funding

Mirella Manfredi was supported by a FAPESP post-doctoral research grant (2015/00553-5). Paulo S. Boggio is sup- ported by a CNPq research grant (311641/2015-6).

Acknowledgements

We are very grateful to Sofia Macarini for her kind support.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders: DSM-III-R* (3rd ed.). Washington, DC: Ameri- can Psychiatric Association.

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language, 52*(1), 36–53. https://doi.org/10.1016/j.jml.2008.02.005.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*(2), 201–212. https://doi.org/10.1016/0278-2626(90)90029-N.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2.

Braeutigam, S., Swithenby, S. J., & Bailey, A. J. (2008). Contextual integration the unusual way: A magnetoencephalographic study of responses to semantic violation in individuals with autism spectrum disorders. *European Journal of Neuroscience, 27* (4), 1026–1036. https://doi.org/10.1111/j.1460-9568.2008.06064.x.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146*(1), 172–184. https://doi.org/10.1016/j.brainres.2006.07.033.

Coderre, E. L. (2020). Dismantling the "visual ease assumption:" A review of visual narrative processing in clinical populations. *Topics in Cognitive Science, 12*(1), 224–255. https://doi.org/10.1111/tops.12446.

Coderre, E. L., Cohn, N., Slipher, S. K., Chernenok, M., Ledoux, K., & Gordon, B. (2018). Brain and language visual and linguistic narrative comprehension in autism spectrum disorders: Neural evidence for modality-independent impairments. *Brain and Language, 186*(June), 44–59. https://doi.org/10.1016/j.bandl.2018.09.001.

Cohn, N. (2012). Explaining 'I Can't Draw': Parallels between the structure and development of language and drawing. *Human Development, 55*(4), 167–192. https://doi.org/10.1159/000341842.

Cohn, N. (2013). Visual narrative structure. *Cognitive Science, 37*(3), 413–452.

Cohn, N. (2019). Your Brain on Comics : A Cognitive Model of Visual Narrative Comprehension. 1–35. https://doi.org/10.1111/tops.12421.

Cohn, N. (2020). Visual narrative comprehension: Universal or not? *Psychonomic Bulletin & Review, 27*, 266–285. https://doi.org/10.3758/s13423-019-01670-1.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea) nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology, 65*(1), 1–38. https://doi.org/10.1016/j.cogpsych.2012.01.003.

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition, 70*(1), 42–52. https://doi.org/10.1016/j.bandc.2008.12.005.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*(3), 494–506. https://doi.org/10.1162/0898929053279568.

Cummings, A., Ceponiene, R., Dick, F., Saygin, A. P., & Townsend, J. (2008). A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research, 1208*, 137–149. https://doi.org/10.1016/j.brainres.2008.02.015.

de Mello, Argollo, N., Shayer, B. P. M., Abreu, N., Godinho, K., Durán, P., et al. (2011). Versão abreviada do WISC-III: correlação entre QI estimado e QI total em crianças brasileiras. *Psicologia: Teoria e Pesquisa, 27*(2), 149–155.

Dunn, M. a., & Bates, J. C. (2005). Developmental Change in Neutral Processing of Words by Children with Autism. Journal of Autism and Developmental Disorders, 35(3), 361–376. https://doi.org/10.1007/s10803-005-3304-3.

Dunn, M., Vaughan, H., Jr., Kreuzer, J., & Kurtzberg, D. (1999). Electrophysiologic correlates of semantic classification in autistic and normal children. *Developmental Neuropsychology, 16*(1), 79–99. https://doi.org/10.1207/S15326942DN160105.

Fishman, I., Yam, A., Bellugi, U., Lincoln, A., & Mills, D. (2011). Contrasting patterns of language-associated brain activity in autism and williams syndrome. *Social Cognitive and Affective Neuroscience, 6*(5), 630–638. https://doi.org/10.1093/scan/nsq075.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89.

Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462.

Kujala, T., Lepistö, T., & Näätänen, R. (2013). The neural basis of aberrant speech in autism spectrum disorders. *Neuroscience & Biobehavioral Reviews, 37*, 1–8. http://www.sciencedirect.com/science/article/pii/S0149763413000079.

Kuperberg, G. R. (2013). The pro-active comprehender: What event- related potentials tell us about the dynamics of reading compre- hension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unrave- ling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 176–192). Baltimore: Paul Brookes Publishing.

Kutas, M, & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207(4427), 203 LP – 205. https://doi.org/10.1126/science.7350657.

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research, 1066*(1–2), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052.

Liu, B., Wang, Z., Wu, G., & Meng, X. (2011). Cognitive integration of asynchronous natural or non-natural auditory and visual information in videos of real-world events: An event-related potential study. *Neuroscience, 180*, 181–190. https://doi.org/10.1016/j.neuroscience.2011.01.066.

Lord, C., Rutter, M., & Couteur, A. (1994). Autism diagnostic inter- view-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

Manfredi, M., Cohn, N., & Kutas, M. (2017). When a hit sounds like a kiss: An electrophysiological exploration of semantic processing in visual narrative. *Brain and Language, 169*. https://doi.org/10.1016/j.bandl.2017.02.001.

Manfredi, M., Cohn, N., De Araújo Andreoli, M., & Boggio, P. S. (2018). Listening beyond seeing: Event-related potentials to audiovisual processing in visual narrative. *Brain and Language, 185*, 1–8. https://doi.org/10.1016/j.bandl.2018.06.008.

Manfredi, M., Cohn, N., Sanchez Mello, P., Fernandez, E., & Boggio, P. S. (2020). Visual and verbal narrative comprehension in children and adolescents with autism spectrum disorders: An ERP study. *Journal of Autism and Developmental Disorders, 0123456789*. https://doi.org/10.1007/s10803-020-04374-x.

Massaro, D. W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal of Developmental and Learning Disorders, 7*(October), 111–144.

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 51*(3), 277–286. https://doi.org/10.1111/j.1469-7610.2009.02157.x.

Mongillo, E. a, Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. Journal of Autism and Developmental Disorders, 38(7), 1349–1358. https://doi.org/10.1007/s10803-007-0521-y.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews, 36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). Online integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 605–616.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia, 48*(10), 2940–2951. https://doi.org/10.1016/j.neuropsychologia.2010.06.003.

Plante, E., Petten, C. V, & Senkfor, a J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. Neuropsychologia, 38(13), 1669–1684. http://www.ncbi.nlm.nih.gov/pubmed/11099725.

Proverbio, A. M., Calbi, M., Manfredi, M., & Zani, A. (2014). Audio-visuomotor processing in the Musician's brain: An ERP study on professional violinists and clarinetists. *Scientific Reports, 4*, 5866. https://doi.org/10.1038/srep05866.

Puce, A., Epling, J. a, Thompson, J. C., & Carrick, O. K. (2007). Neural responses elicited to face motion and vocalization pairings. Neuropsychologia, 45(1), 93–106. https://doi.org/10.1016/j.neuropsychologia.2006.04.017.

Quante, L., Bölte, J., & Zwitserlood, P. (2018). Dissociating predictability , plausibility and possibility of sentence continuations in reading : evidence from late-positivity ERPs. 1–24. https://doi.org/10.7717/peerj.5717.

Ribeiro, T. C., Valasek, C. a, Minati, L., & Boggio, P. S. (2013). Altered semantic integration in autism beyond language: a cross-modal event-related potentials study. Neuroreport, 24(8), 414–418. https://doi.org/10.1097/WNR.0b013e328361315e.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events. 2037–2057.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657–671. https://doi.org/10.1162/089892999563724.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*(3), 701–718. https://doi.org/10.1016/S0926-6410(03)00196-4.

van Petten, C. K., & Rheinfelder, H. (1995). Conceptual Relationships Between Words and Environmental Sounds: Event-related Brain Potentials Measures. *Neuropsychologia, 33*(4), 485–508. http://www.sciencedirect.com/science/article/pii/002839329400133A.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children* (3rd edn rev). NewYork: Psychological Corporation.

Wechsler, D. (1999). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234–245. https://doi.org/10.1016/j.bandl.2006.12.003.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654–667.

Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review, 14*, 57–63.