Neuropsychologia 51 (2013) 1002–1010

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

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

Multilevel alterations in the processing of audio–visual emotion expressions in autism spectrum disorders

Geneviève Charbonneau [a], Armando Bertone [b,c], Franco Lepore [a,d], Marouane Nassim [c], Maryse Lassonde [a,d], Laurent Mottron [c], Olivier Collignon [a,d,e,*]

[a] Centre de Recherche en Neuropsychologie et Cognition (CERNEC), Département de psychologie, Université de Montréal, CP 6128, succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7
[b] School/Applied Child Psychology, Department of Education and Counseling Psychology, McGill University, 3700 McTavish St. Montreal, Québec, Canada H3A 1Y2
[c] Centre d'excellence en Troubles envahissants du développement de l'Université de Montréal (CETEDUM), Hôpital Rivière-des-Prairies, 7070 boul. Perras, Montréal, Québec, Canada H1E 1A4
[d] Centre de Recherche CHU Sainte-Justine, 3175, Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada H3T 1C5
[e] Centre for Mind/Brain Sciences (CIMeC), University of Trento, via delle Regole, 101, Mattarello (TN), Italy.

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form
27 January 2013
Accepted 13 February 2013
Available online 24 February 2013

Keywords:
Autism spectrum disorder
Multisensory
Emotion expressions
Audition
Vision

ABSTRACT

The abilities to recognize and integrate emotions from another person's facial and vocal expressions are fundamental cognitive skills involved in the effective regulation of social interactions. Deficits in such abilities have been suggested as a possible source for certain atypical social behaviors manifested by persons with autism spectrum disorders (ASD). In the present study, we assessed the recognition and integration of emotional expressions in ASD using a validated set of ecological stimuli comprised of dynamic visual and auditory (non-verbal) vocal clips. Autistic participants and typically developing controls (TD) were asked to discriminate between clips depicting expressions of disgust and fear presented either visually, auditorily or audio-visually. The group of autistic participants was less efficient to discriminate emotional expressions across all conditions (unimodal and bimodal). Moreover, they necessitated a higher signal-to-noise ratio for the discrimination of visual or auditory presentations of disgust versus fear expressions. These results suggest an altered sensitivity to emotion expressions in this population that is not modality-specific. In addition, the group of autistic participants benefited from exposure to bimodal information to a lesser extent than did the TD group, indicative of a decreased multisensory gain in this population. These results are the first to compellingly demonstrate joint alterations for both the perception and the integration of multisensory emotion expressions in ASD.

1. Introduction

The ability to recognize emotional expressions is a fundamental cognitive ability for the regulation of interpersonal interactions (Adolph, 2002; Custrini & Feldman, 1989; Izard et al., 2001). The tone of the voice and the facial expression are two crucial cues that we constantly use to predict others' actions and to react appropriately in a social situation. An important aspect of affect perception in everyday life is that it usually involves, like speech, the activation of several sensory channels simultaneously. Therefore, the combination of information from facial expression (visual signal) and prosody (auditory signal) usually results in a unified and more optimal representation of the expressed emotion (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; de Gelder & Vroomen, 2000; de Gelder et al., 2005). For example, it has been shown that the multisensory integration (MSI) of these two types of information typically allows for faster and more accurate recognition of emotion expressions in human observers (Collignon et al., 2008, 2010; de Gelder & Vroomen, 2000; Dolan, Morris, & de Gelder, 2001; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Massaro & Egan, 1996) and in human-machine interfaces (Busso et al., 2004).

Deficits in the perception of emotion expressions have been suggested as possible causes of atypical social and communicative interactions that are a striking part of the behavioral phenotype of autistic spectrum disorders (ASD) (Bachevalier & Loveland, 2006; Monk et al., 2010; Sigman, Dijamco, Gratier, & Rozga, 2004).

* Corresponding author at: CIMeC—Center for Mind/Brain Sciences, University of Trento, via delle Regole, 101, Mattarello (TN), Italy. Tel.: +39 461 282778; fax: +39 461 883066.
E-mail addresses: genevieve.charbonneau.1@umontreal.ca (G. Charbonneau), armando.bertone@mcgill.ca (A. Bertone), franco.lepore@umontreal.ca (F. Lepore), marouane.nassim.hrdp@ssss.gouv.qc.ca (M. Nassim), maryse.lassonde@umontreal.ca (M. Lassonde), mottronl@istar.ca (L. Mottron), olivier.collignon@unitn.it (O. Collignon).

0028-3932/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.009

However, a majority of the empirical investigations in the field have focused on the facial expression of emotions using static stimuli such as photographs (Bal et al., 2010), with only a few studies using video representing dynamic facial movements (Golan, Baron-Cohen, & Golan, 2008; Loveland, Steinberg, Pearson, Mansour, & Reddoch, 2008; Loveland et al., 1997). The former static stimuli have limited ecological validity and neglect the intrinsic dynamic nature of facial expressions. Indeed, facial movements have been shown to enrich emotional expression, contributing to its identification and playing an important role in the perception of its intensity (Ambadar, Schooler, & Cohn, 2005; Biele & Grabowska, 2006). Also, neuroimaging studies have shown that the brain regions involved in the processing of facial affect, such as the posterior superior temporal sulcus (pSTS), the amygdala and the insula, respond differently to dynamic than to static emotional expressions (Haxby, Hoffman, & Gobbini, 2000, 2002; LaBar, Crupain, Voyvodic, & McCarthy, 2003; Miki, Takeshima, Watanabe, Honda, & Kakigi, 2011). Moreover, only a few studies explored the processing of affective vocalizations in ASD (Baker, Montgomery, & Abramson, 2010; Hall, Szechtman, & Nahmias, 2003; Loveland et al., 2008; Wang, Lee, Sigman, & Dapretto, 2007). In most cases, these studies included semantic or lexical confounds in the tasks (Lindner & Rosen, 2006) raising the possibility that the results were influenced by differences in language comprehension (Haviland, Walker-Andrews, Huffman, Toci, & Alton, 1996; Paul, Augustyn, Klin, & Volkmar, 2005). Finally, most studies investigating the recognition of emotions in autistic individuals explored a single sensory modality at a time, whereas in natural settings, emotions are expressed both facially and vocally, allowing the combination of these sources of information by human observers for optimal recognition (Collignon et al., 2008; de Gelder et al., 1999; de Gelder & Vroomen, 2000; de Gelder et al., 2005). The use of multisensory conditions to explore the recognition of emotional expressions in ASD is of particular interest since differences in multisensory processing between ASD and typically developing controls (TD) has recently been demonstrated (Collignon et al., 2012; Magnee, de Gelder, van Engeland, & Kemner, 2007; 2008; Russo et al., 2010; Russo, Mottron, Burack, & Jemel, 2012).

An additional challenge associated with the processing of emotional expressions in natural settings is related to the fact that the saliency of emotional information in faces and voices is often reduced by environmental noise. In signal processing, noise can be considered unwanted data that is not being used to transmit a signal, but is simply a by-product of other activities. For example, the voice of an individual can be masked by noise from other human voices or from objects surrounding him. Similarly, a person's facial expression can be partially hidden by an object or because of the angle in which the observer is positioned. Therefore, the ability of the observer to extract efficiently emotional information from noise appears crucial for effective social interactions and therefore it is relevant to evaluate the perception of emotional expressions in noisy situations (Pelli & Farell, 1999). Some studies have suggested that ASD have a specific difficulty in perceiving speech when presented in a noisy background compared to TD (Alcantara, Weisblatt, Moore, & Bolton, 2004; Smith & Bennetto, 2007). To our knowledge, no study has investigated the perception of visual or auditory emotional expressions in noise in ASD.

The goal of the present study was therefore to explore the perception and the integration of emotion expressions in ASD by using ecological and validated sets of dynamic visual and non-verbal vocal clips of emotional expressions (Belin, Fillion-Bilodeau, & Gosselin, 2008; Simon, Craig, Gosselin, Belin, & Rainville, 2008). Participants were asked to categorize expressions of fear or disgust as quickly and accurately as possible when presented with auditory, visual and audio–visual stimuli. This task allowed us to compare recognition and MSI performance of emotional expressions between ASD and TD. We also compared unisensory performance of ASD and TD participants by measuring their ability to discriminate emotional expressions when presented auditorily and visually in individually adapted levels of noise. Similar paradigms have been previously successfully used to demonstrate that the perception of emotional expressions is a robust multisensory situation which follows rules that have been observed in other perceptual domains (Collignon et al., 2008) and to illustrate gender differences in the processing of emotion expressions (Collignon et al., 2010).

2. Material and methods

2.1. Subjects

Thirty-two autistic participants (30 males; mean age 21 years $\pm$ 6; range 14–32 years) and 18 TD controls (18 males; mean age 21 years $\pm$ 4; range 15–27 years) participated in this study. Participants were recruited from the database of the Rivière-des-Prairies Hospital's autism clinic (Montreal, Canada). ASD participants were defined using DSM-IV-TR diagnostic criteria, as operationalized by the Autism Diagnostic Interview – Revised (ADI-R) (Lord, Rutter, & Le Couteur, 1994) and the Autistic Diagnostic Observation Schedule-Generic (ADOS-G) (Lord et al., 2000) algorithms. Control participants and their first-degree relatives were screened with a questionnaire for any history of neurological or psychiatric disorders. The groups were closely matched in terms of laterality and Weschler IQ ([Full-scale = ASD: 105 $\pm$ 15; TD 111 $\pm$ 9]; [Performance = ASD: 102 $\pm$ 13; TD 108 $\pm$ 10]; [Verbal = ASD: 106 $\pm$ 16; TD 112 $\pm$ 11]). All participants had a global Wechsler score of 80 or more. They all had normal or corrected to normal far and near vision assessed before testing using near and far Snellen acuity charts. The ethics board of both the Rivière-des-Prairies Hospital and University of Montreal approved the study.

Autism lies on a spectrum, and comprises two major subgroups: individuals with classic autism and those with Asperger's syndrome. These groups share the combination of social-communication difficulties, repetitive behaviors and restricted interests. In classic autism, language development in children is also delayed and their intelligence level range from intellectual disability to superior intelligence, while criteria for Asperger excludes cognitive impairment. Following the DSM-5 decision to adopt a dimensional view of heterogeneity in autism spectrum, we subsequently merged the subgroups in a common sample of ASD. However, since previous experiments suggested that classic autism and Asperger's syndrome might differ in terms of perceptual abilities (Brochu-Barbeau, Soulières, Dawson, Zeffiro, & Mottron, 2013; Bonnel et al., 2010; Jones et al., 2009), we also analyzed the data by separating the two subgroups and no significant difference was observed between them (see supplemental analyses; SFigs. 2–5).

2.2. Stimuli

As in our previous experiments, fear and disgust expressions were used because, from an evolutionary perspective, these emotions may be more important for survival than other basic emotions. Indeed, in the multisensory domain, Dolan et al. (2001) suggested that the rapid integration across modalities is not as automatic for happy expressions as it is for fear signals. More specifically, the goal of fear would be to augment sensory vigilance (Davis & Whalen, 2001) whereas disgust is associated with sensory rejection (Rozin & Fallon, 1987). Consistent with this idea, it has been demonstrated that fear enhances sensory acquisition and perception, whereas disgust dampens it (Susskind et al., 2008), therefore giving empirical support to the Darwinian hypothesis that some basic emotion expressions may have originated in altering the sensory interface with the physical world (Darwin, 1972/1998). Furthermore, disgust and fear expressions convey highly discriminable signals (Belin et al., 2008; Ekman & Friesen, 1976; Simon et al., 2008; Susskind et al., 2008) and serve as a model to study the existence of separate neural substrates underlying the processing of individual emotion expressions (Calder, Lawrence, & Young, 2001).

The *visual stimuli* used in this study were selected from a standardized set of dynamic color stimuli of actors and actresses displaying prototypical facial expressions (Simon et al., 2008). One actor and one actress who best depicts facial expressions of fear and disgust based on a previous control study (Collignon et al., 2008) were selected. The facial expressions were "prototypical" and "natural" insofar as they possessed the key features (identified using the Facial Action Coding System: FACS) identified by Ekman and Friesen (1976) as being representative of everyday facial expressions (Simon et al., 2008). The same actor and actress portrayed the two emotions. The selected clips were edited in short segments of 500 msec with a size of 350 $\times$ 430 pixels using Adobe Premiere and