ASHA American Speech-Language-Hearing Association

JSLHR

Review Article

Multichannel Perception of Emotion in Speech, Voice, Facial Expression, and Gesture in Individuals With Autism: A Scoping Review

Minyue Zhang,[a] Yu Chen,[a] Yi Lin,[a] Hongwei Ding,[a] and Yang Zhang[b]

[a] Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, China [b] Department of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, Twin Cities, Minneapolis

ARTICLE INFO

Article History:
Received August 11, 2021
Revision received December 2, 2021
Accepted January 2, 2022

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

https://doi.org/10.1044/2022_JSLHR-21-00438

ABSTRACT

Purpose: Numerous studies have identified individuals with autism spectrum disorder (ASD) with deficits in unichannel emotion perception and multisensory integration. However, only limited research is available on multichannel emotion perception in ASD. The purpose of this review was to seek conceptual clarification, identify knowledge gaps, and suggest directions for future research.
Method: We conducted a scoping review of the literature published between 1989 and 2021, following the 2005 framework of Arksey and O'Malley. Data relating to study characteristics, task characteristics, participant information, and key findings on multichannel processing of emotion in ASD were extracted for the review.
Results: Discrepancies were identified regarding multichannel emotion perception deficits, which are related to participant age, developmental level, and task demand. Findings are largely consistent regarding the facilitation and compensation of congruent multichannel emotional cues and the interference and disruption of incongruent signals. Unlike controls, ASD individuals demonstrate an overreliance on semantics rather than prosody to decode multichannel emotion.
Conclusions: The existing literature on multichannel emotion perception in ASD is limited, dispersed, and disassociated, focusing on a variety of topics with a wide range of methodologies. Further research is necessary to quantitatively examine the impact of methodological choice on performance outcomes. An integrated framework of emotion, language, and cognition is needed to examine the mutual influences between emotion and language as well as the cross-linguistic and cross-cultural differences.
Supplemental Material: https://doi.org/10.23641/asha.19386176

Autism spectrum disorder (ASD) is a range of neurodevelopmental disorders characterized by difficulties with social communication, repetitive behaviors, and restricted interests (American Psychiatric Association [APA], 2013). A considerable number of studies have documented impairments in socio-communicative functioning in the ASD individuals and their constant problems in understanding others' emotional and mental states (Frith & Hill, 2004; Harms et al., 2010; Loukusa et al., 2014). In particular, deficits in the perception of emotional signals are considered as an essential obstacle to their social interactions and relationship development (Kahana-Kalman & Goldman, 2008). Previous studies have identified emotion perception as a significant mediator for the adaptive functioning in individuals with ASD (Hudepohl et al., 2015). Difficulties in emotion perception are particularly prominent in ASD (Baron-Cohen, 1995), and such difficulties are correlated with the low effectiveness with which some autistic individuals perform daily personal and social activities, leading to their lower socialization abilities (Hudepohl et al., 2015). Although researchers have long recognized the importance of emotion cognition in early development for children with or without ASD, there are some terminology confusions and knowledge gaps in the area of multisensory emotion research.

Correspondence to Hongwei Ding: hwding@sjtu.edu.cn. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Emotional cues in speech communication are delivered in several sensory *modalities,* including the auditory modality and the visual modality. Within each modality, emotional information can be conveyed through more than one *channel* in congruent or incongruent forms, such as facial expressions, gestures, and eye gaze in the visual modality (Meeren et al., 2005), and nonverbal emotional prosody and verbal semantics in the auditory modality (Stewart et al., 2013). This set of conceptual definitions lays out the overlaps as well as clear differences between the two terms, *modality* and *channel*. Indeed, some researchers make a fine distinction between multisensory and multichannel processing of emotion and identify distinct channels within a single modality (Lin et al., 2020; Schwartz & Pell, 2012). However, others consider these two as interchangeable and equivalent terms to indicate the multiplex pathways of emotional information (e.g., Gunes & Piccardi, 2007). Obscuring the distinctions between the two terms is prone to put the pathways of emotional cues into a large and undifferentiated group that blurs the subtle similarities and differences among them. Using *modality* and *channel* indiscriminately could lead to inadequate attention to certain pathways of emotional information. For instance, many ASD studies on multisensory/multimodal emotion processing choose facial expressions to represent the visual modality and intonation the auditory modality (e.g., Magnée et al., 2011), but fewer look into gestures and gaze as visual cues, or emotional lexical semantics that can be delivered either in spoken language or as visual text.

To date, multichannel emotion processing in ASD has been much less investigated, compared with the unichannel and multimodal research. Numerous unichannel studies have reported that people with ASD display atypical perceptual processing in visual or auditory emotional tasks (see the works of Uljarevic & Hamilton, 2013, and Zhang et al., 2021, for reviews). ASD individuals are found to struggle with both emotional prosody (Boucher et al., 2000; Doi et al., 2013; Globerson et al., 2015; Martzoukou et al., 2017; Stewart et al., 2013) and emotional semantics (e.g., content words; Beversdorf et al., 1998; Han et al., 2014; Moseley et al., 2015). Within the auditory modality, studies on emotional prosody perception tend to exclude the confounding effects of semantic processing by using pseudo-utterances or nonlinguistic vocal stimuli (e.g., Scheerer et al., 2020). Very few researchers have examined the interaction between emotional prosody and emotional semantics, particularly with respect to how ASD individuals perceive simultaneous auditory emotional cues in multiple channels. This begs questions about ecological validity because in real-world social communication, the lexical and the prosodic channels are intricately linked in natural speech. Similarly, individuals with ASD are known to have problems processing emotional facial expressions (Black et al., 2017; Castelli, 2005), and they also experience difficulties in body language such as pantomime execution/imitation and recognition (Fabbri-Destro et al., 2019; Fridenson-Hayo et al., 2016). However, within the visual modality, there is also a scarcity of multichannel studies examining ASD participants' abilities to handle multiple visual emotional cues simultaneously.

In addition to the unichannel emotion perception problems, multimodal processing deficits are well documented in the ASD literature. ASD individuals were found to have diminished multisensory integration (mainly audiovisual integration), which may produce cascading effects on other core and related ASD symptomatology (see the work of Feldman et al., 2018, for a review). The channels investigated in these studies have been consistently across sensory modalities, with one in the visual modality and the other in the auditory modality. In other words, the existing research has largely bypassed within-modality multichannel processing of emotion, especially in spoken language that carries both nonverbal emotional (prosody) and verbal (lexical semantics) information.

Knowledge Gap in Multichannel Processing

Multichannel processing of emotion needs special attention as it is often obscured in the dominant discourse on unisensory and multisensory processing. Multichannel processing is distinct from unisensory processing since it involves the integration of more than one stream of emotional information. It is not exactly the same as multisensory processing in that the multiple channels can be either across sensory modalities or within one modality.

Research on multichannel processing of emotion can make theoretical contributions. Studies on ASD have shown support for the complex information processing deficit theory that individuals with ASD would exhibit impairments on tasks with high demands for integration of information, despite their intact or even enhanced simple abilities within each domain (Williams et al., 2006). However, the data have largely focused on unisensory/modal neutral information processing with little attention to multichannel emotional information processing issues, especially those involving channels within a single sensory modality.

The redundant nature of simultaneous multichannel emotional cues may lead to an alternative outcome incompatible with the complex information processing deficit theory. Prior studies on typically developing (TD) participants (Collignon et al., 2008; Dolan et al., 2001) and human–machine interfaces (Busso et al., 2004) have discovered that the convergence of emotional information from various signal sources could result in a unified and optimal representation of emotions, which is reflected in

shorter response time and higher accuracy in emotion recognition. One cannot rule out the possibility that individuals with ASD may be able to take advantage of the concerted and mutually reinforcing emotional cues from different channels. This would imply that the presumed deficits in unisensory emotion perception in ASD might not be so evident in multichannel emotion perception. Two major issues remain unresolved here. Are there deficits in multichannel emotion perception in the ASD population? Under what circumstances does the interaction from multiple channels facilitate or hamper emotion perception?

In light of the existing knowledge gaps due to the prevalent cross-modal experimental protocols, a comprehensive review is in need to guide future works in this emerging area of research and practice with autistic individuals. Multichannel processing of emotion provides a testbed and discussion forum on the integration of emotion, language, and cognition, which is crucial for the intervention and exploration on the ASD population. This integrative approach for emotion cognition within and across modalities is in line with the theoretical framework for the Naturalistic Developmental Behavioral Interventions (Schreibman et al., 2015) that provides the guidelines for basic research and evidence-based practice for understanding the mutual influences between language, emotion, and cognitive skills and knowledge. The ultimate goal is to effectively and efficiently improve learning characterized by cross-domain integration of areas where ASD individuals have core difficulties. Insights from such exploration could foster new training approaches for promoting both linguistic and socio-emotional functioning in clinical populations including autism (e.g., Lindquist et al., 2015). Multichannel investigation involving the lexico-semantic channel, the prosodic channel, and other channels of emotional information would be particularly useful to answer important questions on the presence or absence of multichannel advantage during emotion processing and the developmental changes (Paulmann et al., 2009), which helps seek a developmental explanation that unifies the social and broader emotion-related anomalies of ASD (Gaigg, 2012).

Objectives of the Present Scoping Review

To our knowledge, no systematic analysis and summary report have been conducted specifically on multichannel processing of emotion in ASD. An initial search of empirical studies highlighted a paucity of research and a wide range of study types and methodologies. Therefore, a scoping review was considered more appropriate to summarize the diversity of research on this issue, synthesize the current empirical knowledge, and identify gaps for further research (Arksey & O'Malley, 2005). The scoping approach serves to examine the extent and range of research activity and allows the inclusion of a broad range of studies heterogeneous in design and methods (Barnett et al., 2019; Levac et al., 2010). Specifically, we focused on the following key aspects: (a) study characteristics (especially study aims), (b) paradigms and stimuli, and (c) the performance of participants with ASD on multichannel emotion perception and the factors that might influence the performance (e.g., age, developmental and cognitive levels of participants, variations in the experimental protocols or stimuli).

Method

This scoping review was conducted following the methodological framework proposed by Arksey and O'Malley (2005). The framework consists of five key stages: (a) identifying the research question (stated in the introduction); (b) identifying relevant studies; (c) study selection; (d) charting the data; and (e) collating, summarizing, and reporting the results.

Identifying Relevant Studies

The search strategy to identify relevant studies was based on the three-step method advocated by Aromatiari and Riitano (2014). First, a limited search of Google Scholar was performed to obtain a small selection of relevant articles. We extracted from these papers keywords and index terms that were to be included in a more comprehensive search. We then conducted a systematic search for studies published until July 2021 using major electronic databases (Web of Science, MEDLINE, ERIC, PsycINFO, PsycARTICLES, and Psychology and Behavioral Sciences Collection). In this scoping review, we attempted to focus specifically on the multichannel processing of emotion in individuals with ASD. Keywords used for the population were *autism, autism spectrum disorder, ASD,* and *Asperger,* which were combined with terms relating to multichannel emotion perception: *multichannel, multimodal, cross-channel, cross-modal, emotion, affect, facial, vocal, auditory, face, voice, meaning, prosody,* and *(in)congruent*. In the final step, we manually examined the reference lists of relevant review articles and the included studies for additional potential studies.

Study Selection

We considered studies that met the following criteria: (a) included participants formally diagnosed with ASD by a clinical psychologist or psychiatrist as meeting the criteria of *Diagnostic and Statistical Manual of Mental Disorders–Fourth/Fifth Edition* (*DSM-IV/DSM-5*; APA, 1994, 2013), *International Classification of Diseases, Tenth*

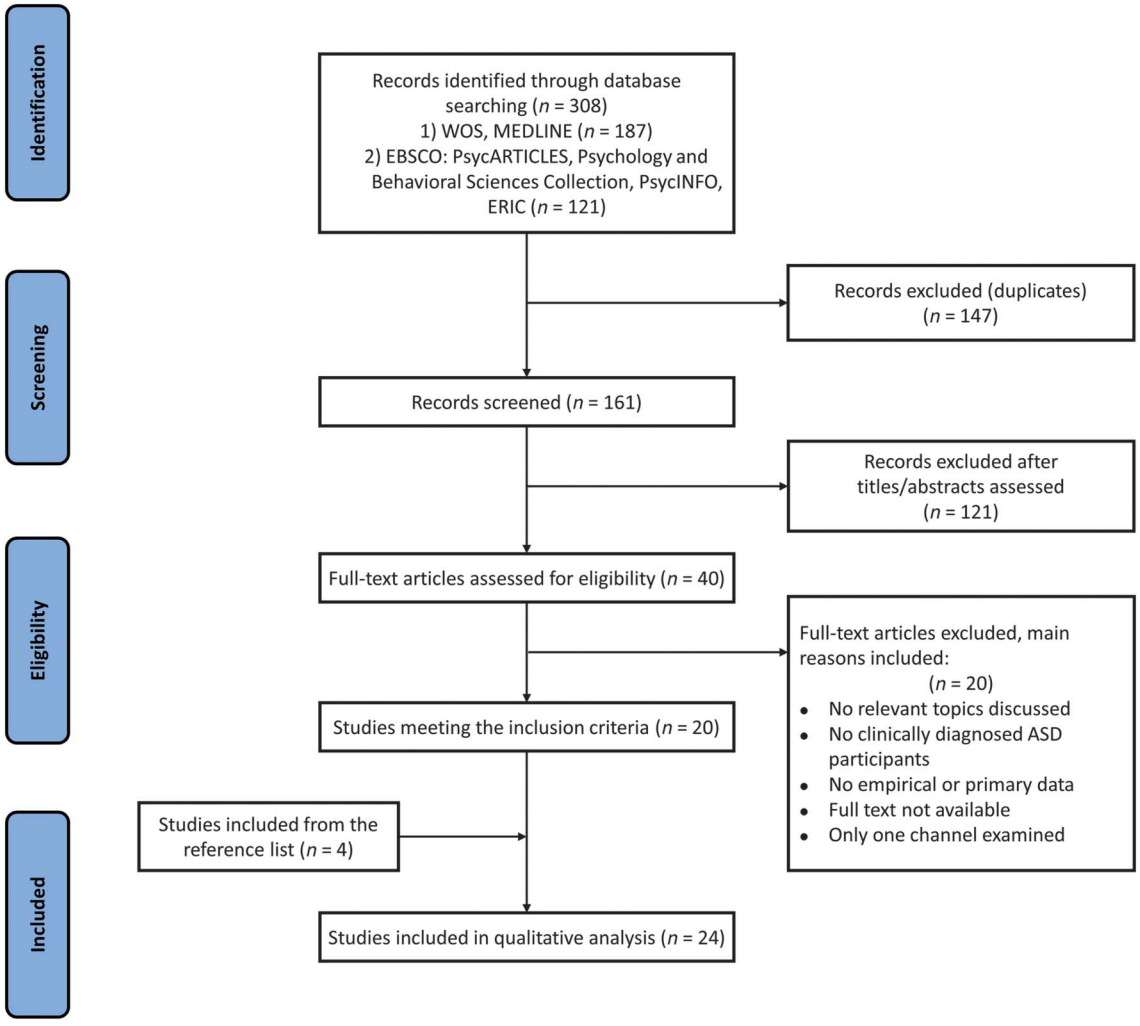
Edition (World Health Organization, 1993), or other valid diagnostic procedures (e.g., *Autism Diagnostic Interview–Revised* [Lord et al., 1994], *Autism Diagnostic Observation Schedule* [Lord et al., 1989], *Asperger Syndrome Diagnostic Scale* [Myles et al., 2001]); (b) reported on the perception of emotional cues in more than one channel (e.g., prosody, lexical meaning, facial expression); (c) were published in English or other languages with an available translation in English; and (d) have full text available. For a scoping review, no restrictions were placed on the age or intelligence of participants because of the limited work on this topic. The inclusion criteria included all multichannel research that employed a variety of study types, designs, or paradigms. Studies were excluded if they were purely theoretical or were secondary research. Publications without clear reporting of aims, methods, or results were also excluded. Opinion papers, comments, or editorials were not included either.

The process of identifying and selecting studies for inclusion is presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram in Figure 1. Our database search identified 308 potentially eligible articles. After duplicates were removed, there were 161 studies left. We then screened the titles, abstracts, and methods against the inclusion and exclusion criteria, resulting in 40 articles for full-text reviews, among which 20 studies met the inclusion criteria. A further four papers were identified through the examination of reference lists, which yielded altogether 24 studies included in this review.

Charting the Data

Data were extracted from each included study in terms of the following key elements: (a) study characteristics (e.g., primary author, publication year, country of

Figure 1. PRISMA flow diagram of study selection. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; ASD = autism spectrum disorder.



study, study aims); (b) task characteristics (e.g., task paradigm, stimulus type); (c) participant information (e.g., ASD sample size, ASD sample demographics, presence or absence of any comparison group); and (d) key findings relating to multichannel processing of emotion in ASD. Where the country of the study was not explicitly stated, we recorded the country of the affiliation of the first author.

Quality Assessment

Two authors independently assessed the quality of each of the included studies using the standard quality assessment (SQA) criteria for evaluating primary research papers from various fields for quantitative studies (Kmet et al., 2004). In the SQA criteria, "quality" was defined with respect to the internal validity of a study, or the extent to which the design, conduct, and analyses minimized errors and biases (Hennekens & Buring, 1987). The checklist for appraising the study quality contains 14 items that examine several crucial aspects of quantitative research, including study objectives, study designs, subject selection methods, subject allocation, controlling, outcome measures, sample sizes, analysis methods, and so forth. These items were scored according to the degree to which the criteria were met ("yes" = 2, "partial" = 1, "no" = 0). Items not applicable to a particular study design, for example, items relating to the use of interventions when assessing a noninterventional study, were marked "n/a" and were excluded from the summary score calculation (i.e., total possible score = [14 – number of excluded items] × 2). A summary score was calculated for each study by summing the total score obtained across relevant items and dividing it by the total possible score. Given that the scoping review aimed to present an overview of existing research regardless of the quality of individual studies, no articles were excluded based on quality criteria (Arksey & O'Malley, 2005; Levac et al., 2010).

Results

Twenty-four studies were eligible for inclusion in the scoping review (see Figure 1 for the description of the selection process). The included articles are detailed in Supplemental Material S1.

Characteristics of the Included Studies

Distribution of Studies by Year and Country of Authorship

Studies included in this review ranged from the years 1989 to 2020. Over half of the included studies were produced in the last 7 years. The number of articles published per year peaked in 2013 and remained relatively stable in the following years. This indicates a trend for increasing and continuous interest in multichannel emotion perception in ASD.

The included studies were from 10 different countries. Five studies were based in the United Kingdom; five in the United States; four in Israel; two in each of France, Japan, and Canada; and one in each of Australia, China, New Zealand, and the Netherlands. The vast majority of the included studies (95.8%) used nontonal languages, and over half of them were conducted with English-speaking populations.

Study Quality

The included studies had a mean quality index score of 0.83 (range: 0.54–0.95, see Supplemental Material S1). Most of the studies (79.2%) attained a score over 0.80, indicating that the overall quality is high. The interrater correlation coefficient (using Spearman correlation; Gwet, 2014) between the two raters was 0.85. Disagreements were resolved by follow-up discussions to reach a consensus.

Study Aims

The 24 studies included in the review mainly sought to achieve four aims. The primary aim of 17 studies is to assess the ability of individuals with ASD to process emotional information through multiple channels compared with TD counterparts (Boucher et al., 2000; Charbonneau et al., 2013; Doi et al., 2013; Golan et al., 2015; Jones et al., 2011; Kahana-Kalman & Goldman, 2008; Lerner et al., 2013; Macdonald et al., 1989; O'Connor, 2007; Philip et al., 2010; Scheerer et al., 2020; Segal et al., 2014; Singh & Harrow, 2014; Su et al., 2018; Taylor et al., 2015; Vannetzel et al., 2011; Xavier et al., 2015).

Two of the included studies focused on examining the interaction between emotional information conveyed through different channels. Specifically, Stewart et al. (2013) aimed to examine the contribution of semantic content to affective prosody recognition as well as the connection between vocal and facial recognition of emotion in the ASD population, and Golan et al. (2018) sought to assess the relative contribution of cues from several perceptual modalities to facial emotion recognition in children with ASD.

Another objective that some of the included studies tried to achieve is to explore the relationship between cognitive mechanisms or abilities and emotion perception in multiple channels in ASD (Globerson et al., 2015; Magnée et al., 2011). Globerson et al. (2015) designed their study to investigate the role played by auditory perceptual abilities and general emotion recognition abilities in affective prosody recognition in ASD, and Magnée et al. (2011) focused on how multichannel processing of emotion in ASD is modulated by attentional capacity.

The last major aim of research in this field is to investigate the efficacy of training or intervention on multichannel emotion processing abilities in ASD (Kandalaft et al., 2013; Lopata et al., 2013; Matsuda & Yamamoto, 2013). Matsuda and Yamamoto (2013) explored whether children with ASD could be taught the relationships between affective prosody and facial expression through cross-channel matching-to-sample training, Lopata et al. (2013) examined the effect of their comprehensive school-based intervention on ASD participants' ability to identify emotions in facial and vocal expressions, and Kandalaft et al. (2013) conducted a feasibility study to investigate the use of virtual reality to enhance verbal and nonverbal emotion recognition and other social skills in young adults on the autism spectrum.

ASD Sample

The sizes of ASD samples varied greatly, ranging from four participants (Matsuda & Yamamoto, 2013) to 99 participants (Jones et al., 2011), with 19 studies (79.2%) using samples between 10 and 30 participants. A total of 10 studies (41.7%) included only "school-aged" children, defined here as children aged 6–12 years, who have reached the age of receiving formal schooling. Eight studies (33.3%) specifically investigated the multichannel processing of emotion in adults. Two studies (8.3%) investigated adolescents aged 13–17 years, and one study focused on "young children" who were 3–5 years old and had not commenced formal schooling. Another three studies included participants in a broader age range. Lerner et al. (2013) included both school-aged children and adolescents, Charbonneau et al. (2013) included adolescents and adults, and Matsuda and Yamamoto (2013) included young children and school-aged children. This reflects the heterogeneity of participants in studies on multichannel emotion perception in ASD.

Methodologies of the Included Studies

Paradigms

Various methodologies were adopted in the studies included, highlighting three major paradigms. One is the "uni-channel emotion recognition task," in which participants were required to recognize emotion conveyed through more than one channel but the emotional information in each channel was presented separately (Doi et al., 2013; Globerson et al., 2015; Golan et al., 2015; Jones et al., 2011; Kandalaft et al., 2013; Lerner et al., 2013; Lopata et al., 2013; Macdonald et al., 1989; Philip et al., 2010; Taylor et al., 2015; Vannetzel et al., 2011). In each condition, only one of the channels (prosody, meaning, and facial expression) contained emotional cues and the other channels, being "neutral," provided no information for emotion recognition. For example, Golan et al. (2015) conducted two tasks of emotion recognition, one for face and one for voice. In the facial emotion recognition task, only human facial expressions conveying emotions were presented, and in the vocal emotion recognition task, stimuli were sentences spoken in emotional intonation but with no emotion-laden semantic content. In both tasks, participants were asked to choose answers that best described the person's feelings from four emotion labels. The effect of the channel was examined through statistical analysis methods, for example, multivariate analysis of variance (MANOVA).

The second kind is the "bi-channel emotion recognition task," which is characterized by presenting emotional information in different channels simultaneously. In some studies that employed this paradigm, the emotions conveyed through different channels were always the same (Charbonneau et al., 2013; Su et al., 2018; Xavier et al., 2015), whereas some other studies used both congruent and incongruent facial–vocal or prosody-meaning stimuli (Magnée et al., 2011; O'Connor, 2007; Segal et al., 2014; Singh & Harrow, 2014; Stewart et al., 2013). In the latter case, none of the studies gave participants any explicit instruction to ignore information conveyed in either channel, but Singh and Harrow (2014) did ask participants to pay attention to the emotional information conveyed through one of the channels.

Another paradigm often adopted by the included studies is the "inter-channel matching task." It was conducted in two forms in the included studies: identification and discrimination. In an interchannel matching (identification) task, participants were required to match emotions across different channels by identifying among several choices the item that portrayed the same emotion as that conveyed in the previously given stimuli (Boucher et al., 2000; Golan et al., 2018; Kahana-Kalman & Goldman, 2008; Matsuda & Yamamoto, 2013; Scheerer et al., 2020). For instance, in the study conducted by Golan et al. (2018), participants were asked to listen to the affective prosody and then identify from three facial expressions the one that displayed the emotion in the affective prosody they had heard. The other form of conducting the matching paradigm is via a discrimination task. O'Connor (2007) used simultaneously presented face and voice expressions as stimuli and asked participants to press the button indicating "same" if the facial expression and the affective prosody portrayed the same emotion and otherwise press the button for "different."

Recent studies have also attempted to combine the above paradigms with other kinds of manipulation to investigate the relation between multichannel processing of emotion in ASD and cognitive factors such as attentional capacity. Magnée et al. (2011) used distractors to manipulate participants' attention to the faces and voices when they were exposed to simultaneously presented congruent or incongruent auditory–visual stimuli.

Stimulus Source

The visual stimuli of the included studies were all from standardized sets of facial expressions, including the Ekman battery of facial emotion (Ekman & Friesen, 1976), the Cambridge Mindreading Face–Voice Battery for Children (CAM-C; Golan et al., 2015), the NimStim set of facial expressions (Tottenham et al., 2009), and so forth. The one used in the largest number of studies was established by Ekman and Friesen (1976), with over 30% of the included studies choosing it as the visual stimulus source. In contrast, half of the included studies (12) used auditory stimuli recorded by recruited professional or semiprofessional actors. Six studies selected auditory stimuli from the standardized set of emotional vocal expressions "the Montreal affective voices" (Belin et al., 2008); two studies from CAM-C (Golan et al., 2015); one study from Diagnostic Analysis of Nonverbal Accuracy-2, which is also a standardized measure of emotion recognition (Nowicki, 2004); and one study from the new Advanced Clinical Solutions for the Wechsler Adult Intelligence Scale–Fourth Edition and the Wechsler Memory Scale–Fourth Edition Social Perception subtest (Pearson, 2009). Another two studies used auditory stimuli extracted from databases or produced by experimenters of previous studies.

Key Findings of the Included Studies

The findings of the included studies mainly address two aspects of multichannel processing of emotion in ASD: the ability to recognize emotions conveyed through multiple channels and the interaction among emotional information via different channels.

Recognition of Multichannel Emotional Information in ASD

The included studies reported results mainly concerning the following questions: whether there is a multichannel emotion recognition deficit in ASD and how well the intervention or training for ASD multichannel emotion processing works.

A vast majority of research in this field has discussed findings regarding the ability to recognize emotions through multiple channels in ASD, but the results seem to be discrepant. Most of the studies discovered multichannel socio-emotional deficits in ASD, suggesting their significant impairments in emotion processing across a range of stimulus channels compared with TD counterparts (Boucher et al., 2000; Charbonneau et al., 2013; Doi et al., 2013; Globerson et al., 2015; Golan et al., 2015, 2018; Lerner et al., 2013; Macdonald et al., 1989; O'Connor, 2007; Philip et al., 2010; Singh & Harrow, 2014; Su et al., 2018; Vannetzel et al., 2011). Participants with ASD were found to perform poorly in visual and auditory emotion recognition tasks (e.g., Charbonneau et al., 2013; Doi et al., 2013; Globerson et al., 2015; Golan et al., 2015) and have difficulty understanding the relationship between affective prosody and facial expressions (e.g., Matsuda & Yamamoto, 2013). They also demonstrated deficiencies in integrating face and voice expressions (e.g., O'Connor, 2007).

However, some investigators found no evidence of impairments in the multisensory multichannel processing of emotion signals in ASD (Jones et al., 2011; Xavier et al., 2015), and a few studies reported intact ability in multichannel processing of emotion under certain circumstances. Kahana-Kalman and Goldman (2008) detected no difference in the interchannel (facial and prosody) matching of emotional expressions between children with autism and TD children when the expressions were portrayed by their mother, but significant differences when portrayed by an unfamiliar woman. Magnée et al. (2011) found that the performance on emotion recognition through multiple channels in ASD is modulated by attentional capacity. They demonstrated that when the attention of participants was directed to the cheeks of the faces in the pictures, similar audiovisual congruency effects were triggered in ASD and TD participants, which indicated that, under this condition, participants with ASD were able to integrate facial and prosodic emotion information. This congruency effect was not observed in the ASD group when their attention was not directed. These two studies suggested that emotion recognition through multiple channels might not be systematically deficient in ASD.

In view of the widely reported poor performance of the ASD population in understanding others' emotions in daily communication, recent years have seen an increase in intervention research on multichannel emotion processing in individuals with ASD (Kandalaft et al., 2013; Lopata et al., 2013; Matsuda & Yamamoto, 2013). A 10-month comprehensive school-based intervention involving face and voice emotion recognition instruction led to a significant increase in ASD participants' ability to identify emotional states in facial and vocal expressions (Lopata et al., 2013), demonstrating the feasibility and efficacy of systematic and comprehensive intervention programs in improving ASD individuals' ability of multichannel emotion recognition. In an intervention study conducted by Matsuda and Yamamoto (2013), it was also found that children with ASD could acquire the generalizable relationship between facial expression and affective prosody through cross-channel matching-to-sample training, which indicates the possibility of promoting cross-channel emotion perception in the ASD population provided that appropriate training is implemented at an early developmental stage.

Interaction Among Emotional Information via Different Channels

Findings on the interaction among emotional information in multiple channels are mostly about the different

impacts that congruent or incongruent emotional cues have on the recognition performance of ASD participants. When participants with ASD are exposed to emotional information provided simultaneously through different channels, whether the information in these channels is congruent or incongruent could have distinct effects on their performance of emotion recognition. Congruent emotional information co-occurring in different channels results in better performance on emotion recognition tasks (Charbonneau et al., 2013; Golan et al., 2015; Stewart et al., 2013; Vannetzel et al., 2011; Xavier et al., 2015). Participants with ASD respond faster and more accurately when congruent emotional signals are presented via facial plus prosodic channels (Charbonneau et al., 2013; Vannetzel et al., 2011; Xavier et al., 2015) or prosodic plus semantic channels (Segal et al., 2014; Singh & Harrow, 2014; Stewart et al., 2013).

When incongruent emotional information co-occurs in different channels, the processing of emotion in individuals with ASD tends to be hampered (Segal et al., 2014; Singh & Harrow, 2014; Stewart et al., 2013). Investigation on this incongruence-induced interference effect in emotion recognition in ASD mainly focuses on the prosodic and semantic channels. The incongruence of emotions conveyed through voice and verbal semantics renders emotion recognition more challenging for individuals with ASD, characterized by reduced response accuracy and prolonged reaction times relative to the age-matched neurotypical controls without ASD (Stewart et al., 2013). The data show evidence of a developmental delay as the judgment of the speaker's emotional intent by the ASD individuals has been found to be similar to TD children at a much younger age (Segal et al., 2014). There are also additional notable aberrant response patterns. Compared with the TD comparison group who gives more weight to the prosodic information when decoding emotion in the incongruent trials, the ASD group tends to deemphasize prosodic cues to emotion and rely more on semantic cues (Scheerer et al., 2020) as they were found to select a semantically matching emotion significantly more than TD controls (Stewart et al., 2013).

Discussion

Main Findings

This scoping review aims to summarize the current state of knowledge and identify gaps in the literature regarding multichannel processing of emotion in ASD. A variety of methodologies have been adopted to explore these questions, which, in essence, consists of three major paradigms: unichannel emotion recognition task, bichannel emotion recognition task, and interchannel matching task. Findings are consistent across studies in answering questions regarding the facilitation and compensation of congruent multichannel emotional cues, the interference of incongruent multichannel emotional signals, and the overreliance on semantics rather than prosody to decode multichannel emotion in ASD, but divergence also exists on the existence of multichannel emotion perception deficits in ASD.

Challenges for the Complex Information Processing Deficit Theory

The complex information processing deficit theory proposes that breakdowns in processing occur in ASD when the information to be handled is inherently complex or becomes complex because of the amount of information or the time constraint (Williams et al., 2006). This breakdown pattern would occur across domains and modalities, leading to selective impairments on tasks with high demands for integration of information even though the ASD individuals may demonstrate relatively spared ability within each domain/modality where the demands placed on the brain's processing capacity are low. While the results of our scoping review partly confirm the theory, they also pose certain challenges for its predictions. The incongruence-induced interference effect in emotion recognition in ASD appears to support this theory, but the congruence-induced facilitation effect does not. The interaction between congruent emotional information conveyed simultaneously through different channels partially compensates for the difficulties that individuals with ASD experience with unichannel emotional information and produces a redundancy gain that significantly facilitates the recognition of emotion in ASD (Charbonneau et al., 2013). This compensatory effect has been observed in the ASD populations on different cognitive levels, including children with AS (Grossman et al., 2000) and lower functioning children with ASD (Xavier et al., 2015). The congruence-induced facilitation results appear to challenge the complex information processing deficit theory about ASD in that increasing the amount of information input and the processing capacity demands does not necessarily hamper across-channel integration or lead to the breakdown in processing multichannel emotional information. The theory–fact incompatibility calls for more theoretical discussion based on empirical evidence that probes into this contradiction and, if necessary, revises the model in terms of uni- and multichannel emotion processing in ASD.

Within- Versus Across-Modality Multichannel Interaction

The findings of the included studies about the interaction between multiple channels within one sensory modality center around the prosodic and the lexical channels, which unanimously show a dominance of the lexical

channel in multichannel emotion perception in ASD. This overreliance on semantic information rather than prosodic cues may contribute to their difficulty in abstracting nonliteral meaning and understanding complex emotions such as irony (Singh & Harrow, 2014; Stewart et al., 2013). In contrast, the included studies on across-modality multichannel processing of emotion discuss very little about the interaction between the visual and the auditory modalities or which modality dominates in emotion perception. Previous literature has suggested a visual dominance for both ASD (Bryson et al., 1990) and TD (Collignon et al., 2008) individuals in multisensory integration, but a recent study (Ross et al., 2021) proposed an auditory dominance in TD children under 8 years old. Therefore, future studies are required to investigate whether across-modality multichannel emotion perception would also show a similar dominance of or overreliance on one channel as within-modality multichannel emotion perception from the perspective of development.

Important Moderator Variables

Age can play an important role in the performance of ASD individuals on multichannel emotion processing, as there might be a developmental change in emotion recognition skills (Segal et al., 2014; Xavier et al., 2015). Emotion perception seems to follow a developmental trajectory from broad valence-based understanding to narrow and more nuanced discrete understanding as children acquire emotion words that serve as a way to bind together the array of multichannel information associated with each discrete emotion category (Shablack & Lindquist, 2019). Their ability to recognize specific discrete emotions may therefore increase with the acquisition of emotional knowledge throughout development. As highlighted in the Results section, ASD participants of the included studies cover a wide age range, from young children who have not started formal schooling to adults over 30 years old. The relatively good performance of ASD adults might thus result from the improvement in emotional processing during the course of development (Charbonneau et al., 2013), with deficits mainly existing during childhood (Magnée et al., 2011).

Beyond the chronological age, the developmental/cognitive level of ASD participants may also moderate multichannel emotion processing. ASD children with better cognitive abilities may experience less difficulty in multichannel emotion processing as better cognitive functioning, both verbal and nonverbal, has been found to correlate with improved emotion recognition abilities (e.g., Jones et al., 2011; Salomone et al., 2019). This link appears to be more characteristic of the ASD group than the TD group (Dyck et al., 2006) as individuals with ASD tend to adopt compensatory mechanisms to aid multichannel emotion perception (Grossman et al., 2000) with the developmental level mediating their ability to employ compensatory strategies (Harms et al., 2010). For example, although individuals with ASD were not limited in perceiving the lexical and prosodic emotional information in the stimuli, their nonverbal IQ scores were found to be a significant predictor of their emotional judgment. Individuals with ASD performed relatively poorly in the incongruent emotional lexical-prosodic condition, as a result of their constrained ability to give priority to prosodic over lexical information for extracting the speaker's intent (Segal et al., 2014). This lexical bias is typically found in much younger TD children, suggesting a developmental delay in ASD individuals with less mature strategies during multichannel emotion recognition.

Variations in the paradigm, the intensity of emotional expressions, and environmental interference could generate different degrees of task demand, which might have led to inconsistent conclusions among existing literature. As indicated in the Results section, a variety of methodologies have been adopted to investigate different aspects of multichannel emotion processing. Variations in the paradigm and task design might substantially affect the outcome. Even the same participant would demonstrate different levels of ability in emotion perception depending on how the tasks are structured (Singh & Harrow, 2014). For example, Boucher et al. (2000) and Scheerer et al. (2020) both found that their ASD participants demonstrated intact ability to extract meaning from prosodic changes but were impaired in prosodic-facial affect matching, probably due to the complexity of the interchannel matching task.

The intensity of emotional expressions could strongly influence the ability to recognize facial expressions and emotional prosody in ASD (Doi et al., 2013). Globerson et al. (2015) and Golan et al. (2018) suggested in their studies that the lack of group differences in emotion recognition performance might result from the high emotional salience of the stimuli. This influence exerted by the intensity of emotional expressions may reflect the tendency of individuals with ASD to employ more "cognitive" strategies instead of "automatic" ones to interpret emotions (Dissanayake et al., 2003; Lerner et al., 2013). When participants with ASD are exposed to stronger visual and auditory emotional signals, their cognitive strategies might be more readily employed, which thereby facilitates emotion processing.

The degree of noisiness in the environment has also been suggested as a possible influential factor for multichannel emotion recognition performance in ASD. The processing of emotional expressions in natural settings is unavoidably influenced by environmental noise. An individual's voice can be masked by other human voices or ambient background sounds, and a person's facial expression can also be partially hidden by objects surrounding

him, which often reduces the saliency of emotional information in voices and faces. ASD participants were found to require a higher signal-to-noise ratio than TD controls for the recognition of emotional expressions in various channels (Charbonneau et al., 2013). Thus, a noisier environment might have caused the ASD participants in some studies to perform particularly poorer as they need to accumulate more evidence before a perceptual decision could be made on the emotional expressions displayed.

Limitations of Existing Studies

Current studies on emotion processing in multiple channels in ASD have adopted various methodologies, some of which might have low validity in assessing multichannel emotion perception in ASD. A considerable number of studies tested emotion recognition in different channels separately and then examined the effect of modality/channel via statistical analysis methods such as MANOVA, correlation, and structural equation modeling, and so forth (e.g., Jones et al., 2011; Lerner et al., 2013; Philip et al., 2010; Vannetzel et al., 2011). This methodology seems to be limited in interpretation, since the results of the statistical analysis can only lead to conclusions about whether performance on emotion perception varied concerning modality or channel, such as those rather cautiously drawn by the authors of these studies. When it comes to disentangling the interaction between different channels, this method might not be as powerful as other methods in which emotional signals through different channels are presented simultaneously to the participants.

The auditory stimuli used in many studies have not been previously validated or standardized. Unlike the visual stimuli, which were all from standardized sets of facial expressions, the auditory stimuli were mostly ad hoc constructed, recorded by actors or volunteers specially recruited for the studies. This requires a validation phase prior to the experiments to demonstrate that these stimuli are qualified for the testing of the research hypotheses, but many studies did not state clearly whether they have included this phase. In some studies, the auditory emotional expressions were even expediently portrayed by the experimenters themselves without any validation process, which is very likely to introduce an experimenter bias (Strickland & Suben, 2012). Nevertheless, this is understandable as auditory emotional expressions usually contain semantic meaning, either neutral or emotion laden, which prevents researchers from selecting auditory stimuli directly from standardized stimulus sets developed in other languages that could not be understood by their participants.

Although the overall quality of the included studies is high (M = 0.83), they are actually mixed in quality (range: 0.54–0.95). Two studies received a quality index score lower than 0.70, both of which are intervention studies. They were rated poorly mainly in terms of randomization and blinding. None of the included intervention studies reported information concerning random allocation to the treatment group or blinding of investigators and participants to intervention, which considerably undermines the reliability of their conclusions about the efficacy of intervention or training. Future intervention or training studies should thus employ larger scale randomized trials with blinding measures appropriately taken and reported to reduce detection and performance bias and avoid overestimated treatment effects.

Most of the relevant studies were conducted in nontonal languages (e.g., English, Dutch, Hebrew, French), with rather insufficient evidence from tonal languages. As two important channels of emotional signals, prosody and semantic meaning have been found to display language- and culture-specific features (Wang et al., 2018; Wierzbicka, 1992). Previous research on auditory and language-related deficits in ASD has also indicated that observation in the ASD population who are native speakers of nontonal languages cannot be adapted directly to those speaking a tonal language (Russo et al., 2008; Yu et al., 2015). As the majority of languages in the world are tonal languages (Yip, 2002), recent findings about domain-specific deficits for lexical tone processing in Chinese children with ASD (Wang et al., 2017; Yu et al., 2015) would naturally raise a question about the potential cross-channel influences of pitch information for linguistic and emotional processing. In this regard, current knowledge on multichannel processing of emotion in ASD is very limited and incomprehensive. More research exploring this issue in native tonal language–speaking individuals with ASD is highly needed to increase diversity in the language samples and address key issues regarding domain specificity of vocal pitch processing deficits in relation to emotion cognition (Schelinski & von Kriegstein, 2019) and phonological processing.

Reviewing the existing studies on multichannel emotion perception in ASD, we found that the channels investigated the most are facial expressions and emotional prosody. By comparison, emotional semantics and body language, and so forth have received scant attention. In addition, the vast majority of the literature focused on the integration across modalities, with very few studies looking into the interaction between channels within each modality. Individuals with ASD have been troubled with sociocommunicative problems, which happen frequently in spoken language that carries both emotional prosody and linguistic content. If research on emotion perception in ASD could take a more integrated view on language, emotion, and cognition and lay more emphasis on the interaction between the lexical channel and the prosodic channel, the findings would surely prove to be even more helpful for ASD research and practical intervention (Schreibman et al., 2015). Such endeavors would also

bring theoretical contribution to tease apart whether the multimodal emotion perception problems are related to the specific deficits in a single modality or the complexity of the stimuli (Scheerer et al., 2020). If the complex information processing deficit theory about ASD does not provide satisfactory explanations for the results of multichannel emotion processing studies, the multimodal integration deficits may thus be attributable more to the modality-specific impairments than the cross-modal integration problem. This would in turn inspire therapists working with ASD individuals to provide congruent emotional information through multiple channels to bring about the facilitation effect on their unimodal and multimodal emotion recognition performance.

Implications and Future Recommendations

Our scoping review revealed a variety of methodologies adopted in studies on multichannel emotion perception in ASD. Additional quantitative research is needed to examine the influence of methodological decisions on the performance of participants with ASD. It would also be beneficial to evaluate the major paradigms often used in this area through empirical experiments or quantitative syntheses to see which paradigm proves to be most powerful in illumination and interpretation concerning the core behavioral symptoms of ASD.

Considering the heterogeneity in the source and the intensity of emotional stimuli that were used in the existing studies, especially in the auditory modality, and the tendency of experimenters to produce the stimuli by themselves without standardization, it is highly necessary for future studies in this area to develop standardized and validated auditory stimuli of emotional expressions, especially in languages that remain comparatively underresearched. Such work would facilitate later research on multichannel emotion perception in ASD and render their findings more theoretically appealing and convincing.

The existing literature has covered both basic and complex emotions, but only a few studies have conducted direct comparisons of basic and complex emotion processing in people with ASD. Quantitative evidence has been reported by empirical studies and meta-analyses about the significant role of emotional complexity in emotion recognition (Fridenson-Hayo et al., 2016; Icht et al., 2021). The ASD population may be comparable to TD controls in terms of basic emotion recognition but shows reduced performance regarding complex emotions, the identification of which might call for more advanced social and language skills that tend to be impaired in this population. The disparity in the findings between basic and complex emotions thus requires future research to invest more effort in examining the basic versus complex emotion differences and whether these differences are suprachannel or specific to certain channels so that intervention programs for ASD individuals could be developed, relying on their intact abilities to process basic emotions through some channels and targeting improvement in processing complex emotions with the aid of multichannel emotional cues.

Given the uneven distribution of studies conducted in different languages and cultures, more research is required to examine multichannel emotion perception in the ASD population with tonal language backgrounds. Further efforts are needed to explore the linguistic and cultural similarities and differences in multichannel emotion perception in ASD and the role that language and cultural backgrounds play in the perception and integration of emotional signals from various channels. In this regard, it is important to study the interaction between linguistic information and emotional information. The key acoustic parameters for emotional prosody include fundamental frequency, loudness, and voice quality (Kotz & Paulmann, 2011; Patel et al., 2011), which may also be employed to convey linguistic as well as other types of paralinguistic information. Some ASD studies did not report deficits in musical emotion perception (Caria et al., 2011; Heaton et al., 2008; Quintin et al., 2011). This might indicate that the emotion perception deficits observed in ASD could be correlated with the impairments in linguistic abilities (Lindquist et al., 2015; Shablack & Lindquist, 2019) in addition to impairments in theory of mind abilities (Fitzpatrick et al., 2018; Rosello et al., 2020).

Conclusions

This scoping review examined 24 relevant studies to provide a synopsis of the current state of knowledge concerning multichannel emotion perception in the ASD population and to identify directions for future research. The review highlighted a variety of methodologies adopted in this underresearched area and showed that factors relating to participants and tasks may explain the divergence of results. Future empirical research and meta-analyses are needed to assess the impact of methodological differences on the experiment outcomes. An integrated view of emotion, language, and cognition is needed to address the role of language in emotion perception and emotional development. As pitch variation is a primary emotional cue in spoken language, special attention is called for the ASD population who are native speakers of tonal languages, which will help us understand the cross-linguistic and cross-cultural aspects of the reported domain-specific deficits of linguistic and affective pitch processing. Standardized stimulus sets of multimodal and multichannel emotional expressions remain to be developed with linguistic diversity in mind to benefit researchers and clinicians working in this area.

Acknowledgments

H. Ding and Y. Zhang were supported by the Major Program of National Social Science Foundation of China (No. 18ZDA293). H. Ding was additionally supported by the Youth Project of Humanities and Social Sciences Foundation of China (No. 18YJC740103).

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Aromatiari, E., & Riitano, D. (2014). Systematic reviews. A guide to the literature search for systematic review. *American Journal of Nursing, 114*(5), 49–56. https://doi.org/10.1097/01.NAJ.0000446779.99522.f6

Barnett, C., Armes, J., & Smith, C. (2019). Speech, language and swallowing impairments in functional neurological disorder: A scoping review. *International Journal of Language & Communication Disorders, 54*(3), 309–320. https://doi.org/10.1111/1460-6984.12448

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press. https://doi.org/10.7551/mitpress/4635.001.0001

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531–539. https://doi.org/10.3758/BRM.40.2.531

Beversdorf, D. Q., Anderson, J. M., Manning, S. E., Anderson, S. L., Nordgren, R. E., Felopulos, G. J., Nadeau, S. E., Heilman, K. M., & Bauman, M. L. (1998). The effect of semantic and emotional context on written recall for verbal language in high functioning adults with autism spectrum disorder. *Journal of Neurology, Neurosurgery, & Psychiatry, 65*(5), 685–692. https://doi.org/10.1136/jnnp.65.5.685

Black, M. H., Chen, N. T. M., Iyer, K. K., Lipp, O. V., Bölte, S., Falkmer, M., Tan, T., & Girdler, S. (2017). Mechanisms of facial emotion recognition in autism spectrum disorders: Insights from eye tracking and electroencephalography. *Neuroscience & Biobehavioral Reviews, 80,* 488–515. https://doi.org/10.1016/j.neubiorev.2017.06.016

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(7), 847–857. https://doi.org/10.1111/1469-7610.00672

Bryson, S. E., Wainwright-Sharp, J. A., & Smith, I. M. (1990). Chapter 21 autism: A developmental spatial neglect syndrome? In J. T. Enns (Ed.), *Advances in psychology* (Vol. 69, pp. 405–427). North-Holland. https://doi.org/10.1016/S0166-4115(08)60468-9

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., & Narayanan, S. (2004). *Analysis of emotion recognition using facial expressions, speech and multimodal information*. Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, United States. https://doi.org/10.1145/1027933.1027968

Caria, A., Venuti, P., & de Falco, S. (2011). Functional and dysfunctional brain circuits underlying emotional processing of music in autism spectrum disorders. *Cerebral Cortex, 21*(12), 2838–2849. https://doi.org/10.1093/cercor/bhr084

Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. *Autism, 9*(4), 428–449. https://doi.org/10.1177/1362361305056082

Charbonneau, G., Bertone, A., Lepore, F., Nassim, M., Lassonde, M., Mottron, L., & Collignon, O. (2013). Multilevel alterations in the processing of audio-visual emotion expressions in autism spectrum disorders. *Neuropsychologia, 51*(5), 1002–1010. https://doi.org/10.1016/j.neuropsychologia.2013.02.009

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242,* 126–135. https://doi.org/10.1016/j.brainres.2008.04.023

Dissanayake, C., Macintosh, K., Repacholi, B., & Slaughter, V. (2003). Mind reading and social functioning in children with autistic disorder and Asperger's disorder. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 213–239). Psychology Press.

Doi, H., Fujisawa, T. X., Kanai, C., Ohta, H., Yokoi, H., Iwanami, A., Kato, N., & Shinohara, K. (2013). Recognition of facial expressions and prosodic cues with graded emotional intensities in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders, 43*(9), 2099–2113. https://doi.org/10.1007/s10803-013-1760-8

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences, 98*(17), 10006−10010. https://doi.org/10.1073/pnas.171288598

Dyck, M. J., Piek, J. P., Hay, D., Smith, L., & Hallmayer, J. (2006). Are abilities abnormally interdependent in children with autism? *Journal of Clinical Child and Adolescent Psychology, 35*(1), 20–33. https://doi.org/10.1207/s15374424jccp3501

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Fabbri-Destro, M., Gizzonio, V., Bazzini, M. C., Cevallos, C., Cheron, G., & Avanzini, P. (2019). The relationship between pantomime execution and recognition across typically developing and autistic children. *Research in Autism Spectrum Disorders, 61,* 22–32. https://doi.org/10.1016/j.rasd.2019.01.008

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 95,* 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Fitzpatrick, P., Frazier, J. A., Cochran, D., Mitchell, T., Coleman, C., & Schmidt, R. C. (2018). Relationship between theory of mind, emotion recognition, and social synchrony in adolescents with and without autism. *Frontiers in Psychology, 9,* 1337. https://doi.org/10.3389/fpsyg.2018.01337

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., Baron-Cohen, S., & Golan, O. (2016). Basic and complex emotion recognition in children with autism: Cross-cultural findings. *Molecular Autism, 7*(1), 52–52. https://doi.org/10.1186/s13229-016-0113-9

Frith, U. E., & Hill, E. E. (2004). *Autism: Mind and brain*. Oxford University Press.

Gaigg, S. (2012). The interplay between emotion and cognition in autism spectrum disorder: Implications for developmental theory. *Frontiers in Integrative Neuroscience, 6,* 113. https://doi.org/10.3389/fnint.2012.00113

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research, 8*(2), 153–163. https://doi.org/10.1002/aur.1432

Golan, O., Gordon, I., Fichman, K., & Keinan, G. (2018). Specific patterns of emotion recognition from faces in children with ASD: Results of a cross-modal matching paradigm. *Journal of Autism and Developmental Disorders, 48*(3), 844–852. https://doi.org/10.1007/s10803-017-3389-5

Golan, O., Sinai-Gavrilov, Y., & Baron-Cohen, S. (2015). The Cambridge Mindreading Face-Voice Battery for Children (CAM-C): Complex emotion recognition in children with and without autism spectrum conditions. *Molecular Autism, 6*(1), 22. https://doi.org/10.1186/s13229-015-0018-z

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *The Journal of Child Psychology and Psychiatry, 41*(3), 369–379. https://doi.org/10.1111/1469-7610.00621

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications, 30*(4), 1334–1345. https://doi.org/10.1016/j.jnca.2006.09.007

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.

Han, D. H., Yoo, H. J., Kim, B. N., McMahon, W., & Renshaw, P. F. (2014). Brain activity of adolescents with high functioning autism in response to emotional words and facial emoticons. *PLOS ONE, 9*(3), Article e91214. https://doi.org/10.1371/journal.pone.0091214

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Heaton, P., Allen, R., Williams, K., Cummins, O., & Happé, F. (2008). Do social and cognitive deficits curtail musical understanding? Evidence from autism and down syndrome. *British Journal of Developmental Psychology, 26*(2), 171–182. https://doi.org/10.1348/026151007X206776

Hennekens, C. H., & Buring, J. E. (1987). *Epidemiology in medicine*. Little, Brown & Co.

Hudepohl, M. B., Robins, D. L., King, T. Z., & Henrich, C. C. (2015). The role of emotion perception in adaptive functioning of people with autism spectrum disorders. *Autism, 19*(1), 107–112. https://doi.org/10.1177/1362361313512725

Icht, M., Zukerman, G., Ben-Itzchak, E., & Ben-David, B. M. (2021). Keep it simple: Identification of basic versus complex emotions in spoken language in individuals with autism spectrum disorder without intellectual disability: A meta-analysis study. *Autism Research, 14*(9), 1948–1964. https://doi.org/10.1002/aur.2551

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., Sauter, D., Tregay, J., Phillips, R. J., Baird, G., Simonoff, E., & Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *The Journal of Child Psychology and Psychiatry, 52*(3), 275–285. https://doi.org/10.1111/j.1469-7610.2010.02328.x

Kahana-Kalman, R., & Goldman, S. (2008). Intermodal matching of emotional expressions in young children with autism. *Research in Autism Spectrum Disorders, 2*(2), 301–310. https://doi.org/10.1016/j.rasd.2007.07.004

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders, 43*(1), 34–44. https://doi.org/10.1007/s10803-012-1544-6

Kmet, L. M., Lee, R. C., & Cook, L. S. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. Alberta Heritage Foundation for Medical Research.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass, 5*(3), 108–125. https://doi.org/10.1111/j.1749-818X.2010.00267.x

Lerner, M. D., McPartland, J. C., & Morris, J. P. (2013). Multimodal emotion processing in autism spectrum disorders: An event-related potential study. *Developmental Cognitive Neuroscience, 3,* 11–21. https://doi.org/10.1016/j.dcn.2012.08.005

Levac, D., Colquhoun, H., & O'Brien, K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(1), 1–9. https://doi.org/10.1186/1748-5908-5-69

Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal Stroop effects. *Journal of Speech, Language, and Hearing Research, 63*(3), 896–912. https://doi.org/10.1044/2020_JSLHR-19-00258

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion. *Current Directions in Psychological Science, 24*(2), 99–108. https://doi.org/10.1177/0963721414553440

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Rodgers, J. D., Smith, R. A., Gullo, G., McDonald, C. A., Mirwis, J., & Toomey, J. A. (2013). Open-trial pilot study of a comprehensive school-based intervention for high-functioning autism spectrum disorders. *Remedial and Special Education, 34*(5), 269–281. https://doi.org/10.1177/0741932512450518

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19,* 185–212. https://doi.org/10.1007/BF02211841

Lord, C., Rutter, M., & Le Couteur, A. (1994). The Autism Diagnostic Interview–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.

Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., & Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: Group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. *International Journal of Language & Communication Disorders, 49*(4), 498–507. https://doi.org/10.1111/1460-6984.12091

Macdonald, H., Rutter, M., Howlin, P., Rios, P., Le Conteur, A., Evered, C., & Folstein, S. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 30*(6), 865–877. https://doi.org/10.1111/j.1469-7610.1989.tb00288.x

Magnée, M. J. C. M., de Gelder, B., van Engeland, H. V., & Kemner, C. (2011). Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials. *PLOS ONE, 6*(8), Article e24196. https://doi.org/10.1371/journal.pone.0024196

Martzoukou, M., Papadopoulou, D., & Kosmidis, M.-H. (2017). The comprehension of syntactic and affective prosody by adults with autism spectrum disorder without accompanying cognitive deficits. *Journal of Psycholinguistic Research, 46*(6), 1573–1595. https://doi.org/10.1007/s10936-017-9500-4

Matsuda, S., & Yamamoto, J. (2013). Intervention for increasing the comprehension of affective prosody in children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*(8), 938–946. https://doi.org/10.1016/j.rasd.2013.04.001

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16518−16523. https://doi.org/10.1073/pnas.0507650102

Moseley, R. L., Shtyrov, Y., Mohr, B., Lombardo, M. V., Baron-Cohen, S., & Pulvermüller, F. (2015). Lost for emotion words: What motor and limbic brain activity reveals about autism and semantic theory. *NeuroImage, 104,* 413–422. https://doi.org/10.1016/j.neuroimage.2014.09.046

Myles, B. S., Bock, S. J., & Simpson, R. L. (2001). Asperger syndrome diagnostic scale. Pro-Ed.

Nowicki, S. (2004). *Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy*. Emory University.

O'Connor, K. (2007). Brief report: Impaired identification of discrepancies between expressive faces and voices in adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 37*(10), 2008–2013. https://doi.org/10.1007/s10803-006-0345-1

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*(1), 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010

Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the multimodal nature of human communication. *Journal of Psychophysiology, 23*(2), 63–76. https://doi.org/10.1027/0269-8803.23.2.63

Pearson, N. C. S. (2009). *Advanced Clinical Solutions for WAIS-IV and WMS-IV: Administration and scoring manual*. The Psychological Corporation.

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., Atkinson, A. P., Calder, A. J., Johnstone, E. C., Lawrie, S. M., & Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine, 40*(11), 1919–1929. https://doi.org/10.1017/s0033291709992364

Quintin, E. M., Bhatara, A., Poissant, H., Fombonne, E., & Levitin, D. J. (2011). Emotion perception in music in high-functioning adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41*(9), 1240–1255. https://doi.org/10.1007/s10803-010-1146-0

Rosello, B., Berenguer, C., Baixauli, I., García, R., & Miranda, A. (2020). Theory of mind profiles in children with autism spectrum disorder: Adaptive/social skills and pragmatic competence. *Frontiers in Psychology, 11,* 2363. https://doi.org/10.3389/fpsyg.2020.567401

Ross, P., Atkins, B., Allison, L., Simpson, H., Duffell, C., Williams, M., & Ermolina, O. (2021). Children cannot ignore what they hear: Incongruent emotional information leads to an auditory dominance in children. *Journal of Experimental Child Psychology, 204,* 105068. https://doi.org/10.1016/j.jecp.2020.105068

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology, 119*(8), 1720–1731. https://doi.org/10.1016/j.clinph.2008.01.108

Salomone, E., Bulgarelli, D., Thommen, E., Rossini, E., & Molina, P. (2019). Role of age and IQ in emotion understanding in autism spectrum disorder: Implications for educational interventions. *European Journal of Special Needs Education, 34*(3), 383–392. https://doi.org/10.1080/08856257.2018.1451292

Scheerer, N. E., Shafai, F., Stevenson, R. A., & Iarocci, G. (2020). Affective prosody perception and the relation to social competence in autistic and typically developing children. *Journal of Abnormal Child Psychology, 48*(7), 965–975. https://doi.org/10.1007/s10802-020-00644-5

Schelinski, S., & von Kriegstein, K. (2019). The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders, 49*(1), 68–82. https://doi.org/10.1007/s10803-018-3681-z

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., Kasari, C., Ingersoll, B., Kaiser, A. P., Bruinsma, Y., McNerney, E., Wetherby, A., & Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(8), 2411–2428. https://doi.org/10.1007/s10803-015-2407-8

Schwartz, R., & Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics. *PLOS ONE, 7*(10), Article e47279. https://doi.org/10.1371/journal.pone.0047279

Segal, O., Kaplan, D., Patael, S., & Kishon-Rabin, L. (2014). Judging emotions in lexical-prosodic congruent and incongruent speech stimuli by adolescents in the autism spectrum. *Folia Phoniatrica et Logopaedica, 66*(1–2), 25–36. https://doi.org/10.1159/000363739

Shablack, H., & Lindquist, K. A. (2019). The role of language in emotional development. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of emotional development* (pp. 451–478). Springer International Publishing. https://doi.org/10.1007/978-3-030-17332-6_18

Singh, L., & Harrow, M. S. (2014). Influences of semantic and prosodic cues on word repetition and categorization in autism. *Journal of Speech, Language, and Hearing Research, 57*(5), 1764–1778. https://doi.org/10.1044/2014_jslhr-l-13-0123

Stewart, M. E., McAdam, C., Ota, M., Peppe, S., & Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism, 17*(1), 6–14. https://doi.org/10.1177/1362361311424572

Strickland, B., & Suben, A. (2012). Experimenter philosophy: The problem of experimenter bias in experimental philosophy. *Review of Philosophy and Psychology, 3*(3), 457–467. https://doi.org/10.1007/s13164-012-0100-9

Su, Q., Chen, F., Li, H., Yan, N., & Wang, L. (2018). *Multimodal emotion perception in children with autism spectrum disorder by eye tracking study*. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia. https://doi.org/10.1109/IECBES.2018.8626642

Taylor, L. J., Maybery, M. T., Grayndler, L., & Whitehouse, A. J. O. (2015). Evidence for shared deficits in identifying emotions from faces and from voices in autism spectrum disorders and specific language impairment. *International Journal of Language & Communication Disorders, 50*(4), 452–466. https://doi.org/10.1111/1460-6984.12146

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions:

Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders, 43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Vannetzel, L., Chaby, L., Cautru, F., Cohen, D., & Plaza, M. (2011). Neutral versus emotional human stimuli processing in children with pervasive developmental disorders not otherwise specified. *Research in Autism Spectrum Disorders, 5*(2), 775–783. https://doi.org/10.1016/j.rasd.2010.09.005

Wang, T., Lee, Y., & Ma, Q. (2018). Within and across-language comparison of vocal emotions in Mandarin and English. *Applied Sciences, 8*(12), 2629. https://doi.org/10.3390/app8122629

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An event-related potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports, 7*(1), 43254–43254. https://doi.org/10.1038/srep43254

Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion, 6*(3–4), 285–319. https://doi.org/10.1080/02699939208411073

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology, 12*(4–5), 279–298. https://doi.org/10.1080/09297040600681190

World Health Organization. (1993). *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic criteria for research*.

Xavier, J., Vignaud, V., Ruggiero, R., Bodeau, N., Cohen, D., & Chaby, L. (2015). A multidimensional approach to the study of emotion recognition in autism spectrum disorders. *Frontiers in Psychology, 6,* 1954. https://doi.org/10.3389/fpsyg.2015.01954

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *Journal of Autism and Developmental Disorders, 45*(11), 3656–3667. https://doi.org/10.1007/s10803-015-2510-x

Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., & Zhang, Y. (2021). Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism,* 136236132199572. https://doi.org/10.1177/1362361321995725