



University of Messina

ISSN 2612-4033

Journal of Clinical & Developmental PsychologyJournal homepage: <http://cab.unime.it/journals/index.php/JCDP/index>**Prosodic Perception and Production in Autism: A Scoping Review**Bonanno M.¹ , Fabiani M.^{1,2} , Cavallini F.^{1,2} , Iacomini S.² ¹ Department of humanities, social sciences and cultural enterprises disciplines, University of Parma, Italy² Tice Cooperativa Sociale, Piacenza, Italy

ABSTRACT

Background: Prosody is the set of suprasegmental elements of language, such as intonation, rhythm, and accent, which play a crucial role in emotional and pragmatic communication.

Methods: We conducted a scoping review to explore prosodic perception and production in individuals with Autism Spectrum Disorder (ASD). This review followed the PRISMA guidelines and selected articles published between 2014 and 2024 through academic databases such as PsycINFO, Scopus, and the Psychology and Behavioural Sciences Collection. We included 14 studies with 440 participants with ASD and 399 neurotypical participants.

Results: The results show considerable heterogeneity: some studies suggest that autistic individuals have difficulty recognizing emotions from prosody, while others show no significant differences compared to neurotypicals. There are also differences in prosodic production, with ASD individuals tending to modulate their voice in atypical ways. The variability of the results could be due to methodological differences, the age of the participants, and the type of stimuli used.

Conclusions: This variability highlights the need for future studies using more standardized protocols to clarify the role of prosody in ASD and its impact on social communication. Clinical implications suggest that professionals should consider prosodic difficulties in assessment and intervention planning, with targeted support to improve social communicative outcomes in individuals with autism.

Keywords: ASD; Prosody; Emotional Communication; Voice modulation; Speech Perception

* Corresponding author: Silvia Iacomini, Tice Cooperativa Sociale, Via Guglielmo De Meis, 25, 29121, Piacenza, Italy

E-mail address: silvia.iacomini@centrotice.it

<https://doi.org/10.13129/2612-4033/0110-4799>

Introduction

Prosody is the aspect of spoken language that emphasizes its rhythmic and melodic characteristics, such as duration, accent, tone, and intonation, which make up the suprasegmental level of speech (Fox, 2000). The prosodic indices include simple phonological elements such as pitch, duration of segments and pauses, intensity, and more complex phonological elements such as accent, rhythm, and intonation. Through these indices, prosody conveys a great deal of information with meaning (Cutler, 1984):

- Intonation, i.e., the variation in the tone and intensity of a statement, affects the understanding of the linguistic message: not only does it allow us to distinguish between interrogative and affirmative sentences, but it also helps us grasp the meaning of a conversation, which can sometimes differ from the straightforward verbal content (e.g., in the case of irony or sarcasm).
- Rhythm, or metrics, refers to variations in spoken language regarding duration, intensity, and tone. It can influence the rhythm of individual words (i.e., words with alternating patterns of accented and unaccented syllables) and the rhythm of sentences (i.e., the different grammatical elements).
- Accent, or 'stress,' concerns the combination of variations in intensity, pitch, and duration of individual syllables, words, and phrases. Thanks to the accent, which emphasizes one word over the others in a sentence, the listener can understand which word is more relevant within the utterance. (Bruno et al. 2022)

Prosody has several functions in communication (Pfanner et al., 2008):

- Syntactic function: for example, the segmentation of utterances. Various prosodic indices, including pauses, indicate the end of an utterance.
- Pragmatic function: This function allows the emphasis of one or another part of the sentence according to the context of reference and thanks to the
- Affective function: Thanks to prosodic variations, it is possible to perceive the emotional state of the interlocutor or, conversely, to express one's emotional state.

To fully understand the meaning of a conversation, it is necessary to combine verbal information with the non-verbal information conveyed by the speaker (Prieto, 2015). Prosody acts as a link between these two types of linguistic information. It includes two components (Prieto et al., 2012): linguistic prosody, which gives spoken language its lexical and syntactic meaning, and affective-emotional prosody, which conveys non-verbal information such as the speaker's intentions, mood, spontaneity, and emotions (Bruno et al., 2022).

One of the pragmatic skills influenced by prosody is emotion recognition (Golan et al., 2007). A recent study by Cowen et al. (2019) showed that prosody can convey up to 12 emotional states, even when the sentence is in a language different from that of the listener. Other research (Mastropieri & Turkewitz, 1999) on using prosodic indices has explored the development of socio-pragmatic skills in children. Even infants show an early sensitivity to the emotions conveyed by prosody in spoken language (Hubscher & Prieto, 2019). This sensitivity, present at birth, is refined in the first months of life through integration with other motor and visual cues. Given the importance of prosodic skills for language development, particularly socio-pragmatic development, it is plausible to hypothesize that abnormalities in these areas may be related to early difficulties in the production and perception of prosody (Bruno et al., 2022; Grossman et al., 2013).

This appears to be the case in autism (ASD), where there are difficulties with prosodic components in both produced and perceived language (Grossman & Tager-Flusberg, 2012; Yankowitz Schultz & Parish-Morris, 2019; Peppé *et al.*, 2011; Lyons et al., 2014)

The way autistic children and adults speak is often unusual or bizarre, and this 'bizarreness,' one of the most apparent clinical signs of the disorder, is partly due to intonation abnormalities. (Patel et al., 2020; Bone et al., 2015; Diehl et al., 2015; Mesibov, 1992; Paul et al., 2005). Prosodic difficulties can occur in both expressive and receptive language: autistic children have difficulty identifying different intonation patterns and reading the emotional meaning of prosody (Pfanner et al., 2008).

These deficits in individuals with ASD have significant implications for their ability to establish social communication. Indeed, although individuals with ASD can express themselves fluently and articulately, they may not be able to engage in interactive communication (Fusaroli et al., 2017; Grossman et al., 2013; Paul et al., 2005). Thus, even if these individuals do not show deficits in verbal intelligence, there may be a deficit in social language (Lewis et al., 2004). For example, they tend to be overly literal in their language understanding and have difficulty understanding humor, irony, and sarcasm (McDonald et al., 2013; Fay & Schuler, 1980).

This review aims to identify studies investigating the perception and production of prosody in individuals with ASD to provide a clearer overall picture of the issue, as the results have been somewhat conflicting.

Although various studies have examined prosody in ASD, the existing literature reveals a considerable degree of methodological heterogeneity, ranging from differences in tasks and stimuli used to sample characteristics to the type of prosodic elements assessed (e.g., linguistic

vs. emotional, perception vs. production). This variability complicates cross-study comparisons and may underlie the inconsistencies found in prior findings. By analyzing these methodological differences, this review aims to clarify the current state of research and offer a more integrated understanding of how prosody is affected in ASD.

Method

The following scoping review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2015).

To select the articles for this review, we consulted several databases available to students enrolled at the University of Parma, particularly PsycINFO, Psychology and Behavioural Sciences Collection, and Scopus. We selected some additional articles through a manual search.

Eligibility criteria

The eligibility criteria used for this review concerned:

- **Year of publication:** We included articles published between 2014 and 2024 and excluded any papers published outside this time frame. This selection was motivated by the intention to include recent studies that reflect the most recent developments in research on prosody in individuals with ASD. Additionally, this ten-year range allows for a sufficiently broad yet up-to-date overview of emerging trends in the scientific literature on this topic.
- **Participants:** We included studies with autistic participants (i.e., children, adolescents, and adults) and neurotypical controls. We included the study by Lui et al. (2023), although it did not include participants with an official diagnosis of ASD, and there was no control group. This study is relevant to the purpose of this review because it included a sample of participants with high AQ scores, which indicate significant autistic traits. Given that the objective of this review was to explore prosodic perception and production across the broader autism spectrum, we considered the inclusion of this study to be valuable in enriching the scope of the review itself. It provides an additional analytical layer by examining prosodic features in individuals who, while not formally diagnosed, exhibit relevant traits associated with the spectrum. We excluded studies that focused on other neurodevelopmental conditions (e.g., ADHD, intellectual disability, specific language impairment).
- **Language:** The selected articles are all in English. We excluded all articles published in languages other than English.

- Article type and study design: This review includes only original peer-reviewed research articles. We included experimental and observational studies with behavioral or neurophysiological methodologies as long as they assessed prosodic perception and/or production. We excluded purely theoretical papers, dissertations, books, and reviews from the analysis.

Search and selection of the studies

The search used the following keywords: Autism, Prosody, Perception, and Voice, combined with the Boolean operator 'AND'.

We initially identified 73 articles. Of these, 26 were removed as duplicates, while we screened the remaining 47 based on the title and abstract. After this screening, we excluded 22 articles as they did not address the research question. Finally, we assessed the full text of the remaining 25 articles, resulting in 14 articles in this review. The article selection process was also reported in a flowchart at the end of the article (see Figure 1). Table 1 lists the characteristics of the studies in the scoping review to facilitate consultation of their primary data. The characteristics reported include authors and year of publication, characteristics of participants, task type, methods, and results. Table 2 highlights the statistical outcomes of the included studies. As this study is a scoping review, no ethical approval was required. All included studies reported obtaining appropriate ethical clearances from their respective institutional review boards or ethics committees, where applicable.

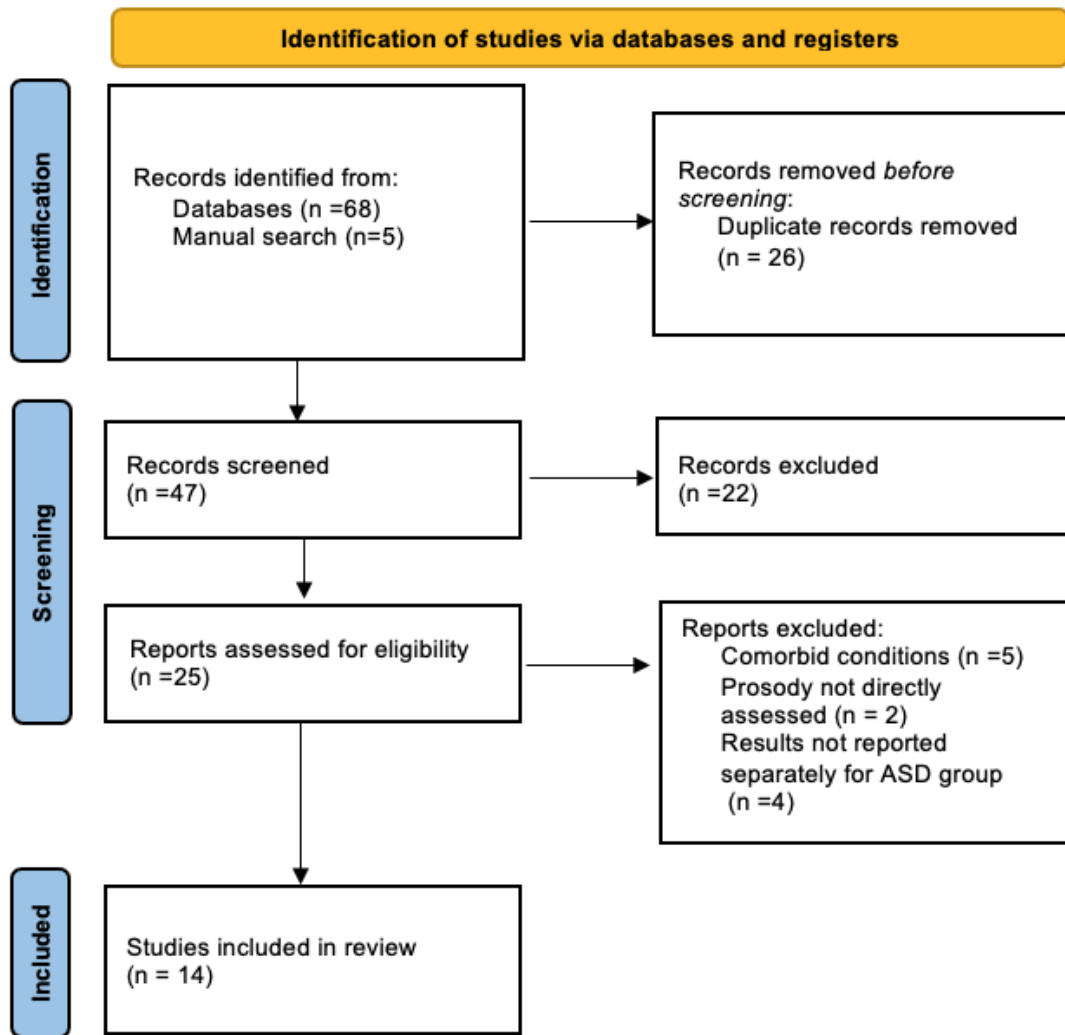


Figure 1. Flowchart describing document selection steps

Authors & Year	Participants	Task type	Method	Results
Ben-David et al., 2019	20 ASD participants Age: 19-27 Gender: F =2; M=18 20 NT participants Age: 22-29 Gender: F =2; M=18	Perception (emotion rating)	T-RES (consisting of 3 tasks: prosodic evaluation, lexical evaluation, general evaluation).	The groups did not differ significantly in their ability to identify emotions. However, for the NT group alone, emotion identification was higher in the prosodic channel than in the the lexical one. About selective attention selective attention, there were no significant differences between the two groups. Finally, the results show that the ASD group assigns relative weights to the prosodic and lexical channels similarly to the NT group.
Chiew et al., 2017	13 ASD participants Age: M= 11.7 Gender: F =1; M=12 21 NT participants Age: M= 8 Gender: F=13, M=8	Perception (emotion recognition)	Recognition of emotion in 63 low-pass filtered sentences	The ASD group recognized the emotions less accurately than the TD group, particularly regarding the neutral and happy emotions. he angry prosody was the one most difficult to recognize in both groups.
Gebauer et al., 2014	19 ASD adults Gender: F=2; M=17 20 NT adults Gender: F=2; M=18	Perception (emotion rating)	Scanned fMRI while listening to 60 semantically unemotional sentences spoken with different prosodies. Used an MR-compatible trackball for emotion ratings on a screen that displayed a visual analogue scale ranging from very sad (-100) to neutral (0) to very happy (+100).	ASD and NT activated the same fronto- temporal and subcortical regions in response to affective prosody; individuals with ASD rated emotional prosody as less intense than NTs.
Gibson et al., 2023	Study 1: 6 ASD and 6 NT speakers 53 NT listeners Age: 18-70 Gender: F=24; M=29	Production and perception	Each of the 12 speakers uttered 7 sentences for each emotional category. Study 1: recognition of the emotion expressed by the speaker. Study 2:	NT speakers generally recognize emotional expressions conveyed more accurately by NT listeners than those expressed by individuals with autism. Concerning vocal modulation, ASD speakers modulate their voice more than NTs to express fear whereas they modulate less for happiness and surprise.

	Study 2: 24 NT listeners Age: 18-22 Gender: F=11; M=13		assessment of valence, vocal modulation and adequacy of voice control on a 7-point likert scale.	
Globerson et al., 2015	23 ASD participants Age: 20-40 Gender: M= 23 32 NT participants Age: 23-39 Gender: M= 32	Perception (emotion recognition)	Psychoacoustic tasks (tone direction recognition and tone discrimination); Prosody tasks (vocal emotion recognition and focus perception); Facial emotion recognition tasks.	In the psychoacoustic tasks, there were no significant differences between the ASD and TD groups; in the prosody tasks the ASD group scored worse in the vocal emotion recognition task, but there were no significant differences in the focus task; in facial emotion recognition no significant differences were found between the ASD and TD groups.
Hoffmann et al., 2023	18 ASD participants Age: 22-62 Gender: F=4; M=14 18 NT participants Age: 23-57 Gender: F=4; M=14	Perception (emotion recognition)	Recognition of emotion with stimuli presented in different modalities (visual, audio-visual, auditory).	Overall, accuracy rates were higher in the NT than in the ASD, despite differences due to the different presentation of the stimuli (both groups had greater accuracy for bimodal stimuli compared to those visual-only or auditory-only stimuli).
Hubbard et al., 2017	<i>Task 1</i> 15 ASD participants Age: 21-42 Gender: M=15 15 NT participants Age: 18-26 Gender: M=15 <i>Task 2</i> 30 NT listeners Age: 18-50 Gender: F= 20; M=10	Production and perception	Task 1: prosodic production task. Task 2: assessment of the emotional context of the sentence and evaluation of the level of naturalness of the emotional content in the sentence on a scale from 0 to 10.	Participants with ASD produced sentences with greater variability of tone, greater intensity, and longer duration than NT participants. Despite this, the listeners identified the sentences produced by ASD speakers more accurately, demonstrating that even if the voice is atypical, it still contains prosodic information.
Leung et al., 2023	38 ASD individuals 38 NT individuals In each group: Age:7-11 (N=14); 12-15 (N=11); 16-56 (N=13) Gender: N/A	Perception (emotion recognition)	Forced-choice emotion labelling task; Navon task; Tone direction discrimination task.	Autistic individuals were as accurate as NT individuals in recognizing emotions from human faces, face-like objects, speech prosody, and from songs.

Lui et al., 2023	94 ASD participants Age: 18-23 Gender: N/A	traits	Perception (Emotion recognition)	Affective prosody recognition tasks; Tone direction recognition; Rapid auditory processing; Social attribution tasks.	The results show that higher AQ scores were correlated with lower accuracy in the affective prosody recognition task for words spoken with low emotional intensity.
Lyons et al., 2014	87 ASD participants Age: 9-17 43 NT participants Age: 13-17 Gender: N/A		Production and perception	Prosody Protocol (PP) Containing 8 subtests (4 perception and 4 production) testing 4 functions that can be mediated by prosody: grammar accent, phraseology and affectation.	Regarding the emotional component of the prosody, no differences emerged between the ASD group and the NT group (neither with regard to the production of emotional prosody production nor for perception). Differences significant differences emerged when analyzing the components of grammar, stress, and phraseology.
Martzoukou et al., 2017	20 ASD participants Age: M=29.8 Gender: F=5; M=15 20 NT participants Age: M=29.1 Gender: F=5; M=15		Perception (emotion recognition)	Part 1: completion of sentences in the shortest possible time; Part 2: recognition of the speaker's emotion	Participants with ASD were less accurate in recognizing emotions than the control group. This difference, however, related to the emotion of surprise. Furthermore, the results show that both groups understood fear and sadness (in higher percentage by the NT group). Similarly, the emotion of happiness was confused with that of surprise (with a higher percentage from the ASD group).
Maslennikova et al., 2022	30 ASD participants Age: 4-6 Gender: F=10; M=20 24 NT participants Age: 4-6 Gender: F=7; M=17		Perception (emotion recognition)	Stimulus administration (i.e. a combination of Russian syllables pronounced with the intonations of joy, anger, sadness, fear and neutral) during an EEG recording and subsequent emotion identification using the following scales: Calm (0-9), Sadness (0-9), Anger (0-9) and Fear (0-9)	There is a significant difference in the perception of emotional prosody by children with ASD compared to NT children in both EEG data and subjective ratings. Also emerged a reduced distinction between EEG responses to joyful and sad prosodies in autistic children compared to the NT group. Children with ASD tended to estimate the sadness of stimuli at a level almost equal to joyful, neutral, and sad.

Schelinski et al., 2019	16 ASD participants Age: 20-51 Gender: F=3; M= 16 16 NT participants Age: 18-52 Gender: F=3; M=13	Perception	Recognition of emotion in 134 semantically neutral nouns. Vocal tone and timbre discrimination test. Montreal Battery of Evaluation of Amusia (online version) including two tests on perception of musical pitch perception of musical pitch: 'scale' and 'out key'.	In recognizing emotions the ASD group performed significantly worse than the NT group for the emotions sadness and fear. Furthermore, the ASD group showed abilities in vocal tone discrimination impaired compared to the NT group. However, there were no significant group differences in tests on vocal timbre discrimination vocal timbre and perception of non-vocal tone.
Wang & Tsao, 2015	25 ASD participants 25 NT participants Age: 6 -11 Gender: N/A	Perception	Completion of the children's communication checklist by parents. Administration of WISC-IV and CLDSC to participants and subsequent emotion-prosodic identification task.	Children with ASD showed a selective deficit in perceiving specific emotions. When presented with prosodies of emotionally neutral and emotionally relevant expressions, children with ASD behaved similarly to NT children in recognizing prosodic signals of negative emotions (i.e., sad and angry) but had more difficulty in recognizing positive emotions (i.e., happy).

Table 1- Characteristics of the selected articles

Study	Statistical outcomes
Ben-David et al., 2019	No significant group differences found in emotion identification ($p = .25$), selective attention ($p = .17$), or integration tasks ($p = .14$). Both ASD and NT groups showed a prosodic dominance pattern ($p < .001$, $\eta^2 = .94$), rating congruent > prosodic > lexical > absent trials. Selective attention failed equally in both groups, with stronger influence of prosody over lexical ($p < .001$, $\eta^2 = .58$).
Chiew et al., 2017	ASD group showed significantly lower accuracy than NT across all prosodic conditions ($p < .001$, $\eta^2 = .450$). Angry prosody was hardest to recognize for both groups ($p = .023$). ASD children took significantly longer to respond ($p = .003$, $\eta^2 = .244$). Nonverbal cognitive ability predicted recognition of neutral and happy prosody ($p = .013$ – $.061$).
Gebauer et al., 2014	ASD rated both sadness and happiness prosody as less intense ($p = .01$, $.02$); NT > ASD activation in left STG; ASD > NT in right caudate at $p < .001$ uncorrected. No group \times emotion interaction.
Gibson et al., 2023	Significant 3-way interaction: Speaker Type \times Sex \times Emotion ($p < .001$, $\eta^2 = .087$). Male ASD speakers were less accurately recognized across all emotions ($p < .001$), while female ASD speakers differed only for fear ($p = .028$). ASD speakers differed from NTs in perceived valence, modulation, and control ($p < .001$). Male ASD speakers rated as less vocally controlled than NTs across emotions; female ASD speakers rated more appropriately for surprise and neutral ($p < .05$).
Globerson et al., 2015	ASD group showed significantly lower accuracy in affective prosody recognition ($p < .01$, $\eta^2 = .14$), but not in pragmatic prosody. No group differences in pitch discrimination or direction. Correlations between pitch perception and prosody were stronger in ASD. Facial emotion recognition predicted vocal emotion recognition only in ASD group ($r = .507$, $p < .05$).
Hoffmann et al., 2023	ASD group showed significantly lower accuracy and slower RTs across all modalities (A/V/AV) with large effect sizes ($d = 1.04$ – 1.17 , $p \leq .002$). Multisensory facilitation of response time was present in NT ($p = .021$) but absent in ASD ($p = .399$); facilitation of accuracy was comparable between groups ($p = .270$).
Hubbard et al., 2017	ASD speakers produced emotional phrases with significantly greater f_0 range ($p < .05$), higher intensity ($p < .05$), and longer durations ($p = .051$ – $.071$) than NT. Perception test: phrases by ASD speakers were more accurately identified (56%) than NT ones (48%) ($p < .001$), but rated as less natural ($p < .001$). Interactions found for speaker \times listener group and emotion context.
Leung et al., 2023	No significant group differences in overall emotion recognition accuracy ($p = .489$); however, a significant interaction emerged between diagnosis, stimulus type, and emotion ($p = .012$). ASD group showed significantly slower response times across all conditions ($p = .002$). Pitch direction discrimination correlated with emotion recognition only in the NT group.
Lui et al., 2023	AQ and RAP significantly predicted accuracy for low-intensity prosody ($p = .019$ and $p = .043$); males performed worse ($p = .015$). A significant AQ \times RAP interaction ($p = .011$, $\eta^2 = .075$) showed that RAP predicted higher APR only in high-AQ individuals. No effects for high-intensity stimuli.
Lyons et al., 2014	Significant group differences for HFA-low vs NT in global perception ($d = 1.21$) and production ($d = 1.02$) in preteens. No differences for HFA-high vs NT. For teens, HFA-low scored lower on production ($d = 1.30$) and global prosody ($d = 1.08$); perception scores were similar across groups. Regression showed language predicted perception ($R^2 = .17$) and nonverbal IQ predicted production ($R^2 = .11$) in preteens; social skills predicted both in teens ($R^2 = .42$ for production).
Martzoukou et al., 2017	ASD group showed significantly lower accuracy than controls for subject prosody (66% vs. 87%, $p = .010$), but not for object prosody (both 88%). RTs were significantly longer in ASD for both conditions ($p < .001$). Affective prosody: ASD group performed significantly worse

	only for surprise (65% vs. 96%, $p < .001$); no differences for other emotions. Syntactic and affective prosody scores were significantly correlated in ASD group ($r = .531$, $p = .016$).
Maslennikova et al., 2022	NT group showed significantly greater EEG PSD differences between joyful, sad, and neutral prosodies than ASD group ($p < .001$, $\eta^2 = 0.58$). In NT children, joy elicited stronger alpha and beta PSD changes ($p = .009-.0006$). ASD group showed reduced hemispheric asymmetry and no EEG-sadness rating correlation.
Schelinski et al., 2019	ASD group showed lower accuracy than NT for sadness and fear ($p = .001$, $d = 1.263$; $p = .005$, $d = 1.061$). Total accuracy: ASD = 68%, NT = 84% ($p = .002$). Pitch discrimination lower in ASD ($p = .033$). Significant correlations only in NT group.
Wang & Tsao, 2015	HFA group showed significantly lower accuracy than NT in identifying happy prosody in both neutral and relevant utterances ($p = .024-.010$). No group differences for sad or angry prosody. Accuracy for happy prosody positively correlated with pragmatic ability ($r = .30-.39$, $p < .05$) and social adaptation ($r = .32-.42$, $p < .05$). Prosody errors correlated with autism severity (CAST score).

Table 2- Statistical outcomes of the studies included

Results

The results are presented below according to four thematic categories in order to explore the heterogeneity of findings and methodological approaches across the included studies (see Table 3).

Perception vs Production

Most studies focused on prosodic perception, with only a few addressing production and even fewer examining both dimensions. Some studies (Gebauer et al., 2014; Maslennikova et al., 2022; Leung et al., 2023) investigated perception through behavioral or neurophysiological paradigms. Their findings varied. Some reported no group differences in recognition accuracy, while others observed significant impairments in the ASD group (see Table 1).

In contrast, studies analyzing prosodic production (e.g., Hubbard et al., 2017; Gibson et al., 2023) showed differences in acoustic features such as pitch range and intensity modulation. Lyons et al. (2014) employed a standardized protocol assessing perception and production, reporting comparable performance in affective prosody but reduced performance in grammatical tasks among participants with ASD.

Age-related differences

The total number of participants corresponds to 440 ASD and 399 neurotypical. The ASD group includes 260 adults aged 16 to 62, 48 adolescents aged 12 to 17, and 132 children aged 4 to 12, while the NT control group includes 261 adults aged 16 to 70, 32 adolescents aged 12 to 17, and 106 children aged 4 to 12.

Participant age emerged as a major moderating factor. Studies involving children (e.g., Maslennikova et al., 2022; Wang & Tsao, 2015; Chiew et al., 2017) reported more pronounced difficulties in affective prosody perception, particularly for happiness and neutral expressions. In contrast, adult-focused studies (e.g., Schelinski & von Kriegstein, 2019; Globerson et al., 2015) identified emotion-specific impairments, particularly in recognizing sadness and fear. Studies involving mixed or broad age ranges (e.g., Leung et al., 2023; Lyons et al., 2014) highlighted differences in reaction time, suggesting that perceptual effort may increase with age in individuals with ASD.

Methodology and task type

Two of the 14 studies used functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) (Gebauer et al., 2014; Maslennikova et al., 2022). On the other hand, the study by Hubbard et al. (2017) involved the participants recording sentences. Finally, all the other studies shared a methodology characterized by the administration of stimuli (i.e., recorded sentences) to be evaluated based on the task to be completed by the participants. Studies using filtered or semantically neutral speech (Gebauer et al., 2014; Chiew et al., 2017) revealed subtler processing differences. Instead, studies using multimodal stimuli (e.g., Hoffman et al., 2023) found more significant impairments. Stimuli also varied in ecological validity. Naturalistic sentences (Wang & Tsao, 2015) elicited different patterns than isolated syllables (Maslennikova et al., 2022).

Labeling tasks (Leung et al., 2023) showed preserved performance in autistic participants, while rating-based paradigms (Gibson et al., 2023) and EEG/fMRI studies (Gebauer et al., 2014; Maslennikova et al., 2022) showed differences not always visible in behavioral accuracy.

Vocal tone discrimination and emotion recognition

Some studies (Leung et al., 2023; Globerson et al., 2015; Schelinski et al., 2019) reported that better discrimination of pitch direction and vocal tone correlated with higher accuracy in emotion recognition in neurotypical participants. However, this correlation was absent or significantly weaker in individuals with ASD. This result suggests a dissociation between low-level acoustic discrimination and high-level affective interpretation in this group.

Some studies showed that individuals with ASD were as accurate as NT individuals in recognizing emotions based on prosody (Gebauer et al., 2014; Leung et al., 2023; Lyons et al., 2014; Ben-David et al., 2019). Other studies (Maslennikova et al., 2022; Hoffmann et al., 2023; Globerson et al., 2015; Chiew et al., 2017) showed a significant difference in emotion recognition ability. Specifically, these studies highlighted that ASD individuals were less accurate than NT ones in recognizing emotions from prosody. Consistent with this, Lui et al.

(2023) showed that a higher AQ score might correlate with lower accuracy in emotion recognition. Finally, other studies (Wang et al., 2015; Martzoukou et al., 2017; Schelinski et al., 2019) showed emotion-specific differences in the perception of emotional prosody between ASD and neurotypical groups.

Vocal modulation and prosodic production

Findings indicated differences in vocal modulation patterns between autistic and neurotypical speakers. Gibson et al. (2023) showed that ASD speakers exhibited a more excellent vocal modulation when expressing fear, whereas NT speakers modulated more effectively for happiness and surprise. Interestingly, there were no differences between NT and ASD speakers for sadness, anger, and neutrality. Female ASD speakers used their voices more 'appropriately' than NT speakers when expressing neutrality and surprise. Male ASD speakers, on the other hand, were consistently perceived as less vocally controlled across all emotions.

Other studies (Gibson et al., 2023; Hubbard et al., 2017) showed that ASD speakers tended to produce sentences with greater intensity and longer duration than their neurotypical counterparts. The Lyons et al. (2014) study revealed that NT individuals were more adept at identifying and producing affirmative and interrogative sentences than ASD ones. Lyons et al. (2014) and Globerson et al. (2015) found no significant differences in the perception or production of speech stress.

Thematic Area		Key Findings	Studies
Perception vs Production		Most studies focused on perception; Production studies showed differences in pitch, intensity, and sentence duration. Few studies addressed both dimensions.	Gebauer et al., 2014; Maslennikova et al., 2022; Leung et al., 2023; Hubbard et al., 2017; Gibson et al., 2023; Lyons et al., 2014
Age-related Differences		Children showed greater deficits, especially for positive/neutral prosody. Adults had emotion-specific impairments (sadness, fear). Mixed-age studies found slower RTs but similar accuracy.	Maslennikova et al., 2022; Wang & Tsao, 2015; Chiew et al., 2017; Schelinski & von Kriegstein, 2019; Leung et al., 2023; Lyons et al., 2014
Methodology and Task Type		fMRI and EEG studies revealed neural-level differences not always reflected in behavior. Multimodal and naturalistic stimuli revealed more impairments. Labeling tasks showed preserved accuracy; rating and neurophysiological tasks highlighted deeper deficits.	Gebauer et al., 2014; Maslennikova et al., 2022; Hoffman et al., 2023; Chiew et al., 2017; Leung et al., 2023; Gibson et al., 2023
Vocal Discrimination and Emotion Recognition	Tone	Pitch discrimination was linked to emotion recognition in NT but not in ASD. Studies showed mixed results: some found no group differences, others found significant impairments or emotion-specific effects. Higher AQ scores predicted lower performance.	Leung et al., 2023; Globerson et al., 2015; Schelinski et al., 2019; Gebauer et al., 2014; Ben-David et al., 2019; Lui et al., 2023; Wang & Tsao,

			2015; Martzoukou et al., 2017
Vocal and Production	Modulation and Prosodic	ASD speakers showed greater modulation for fear, NT speakers better for happiness/surprise. Female ASD speakers rated more appropriate for some emotions. Longer duration and higher intensity in ASD speech. Mixed results in grammatical prosody.	Gibson et al., 2023; Hubbard et al., 2017; Lyons et al., 2014; Globerson et al., 2015

Table 3. Results' summary by thematic areas

Discussion

This scoping review aimed to identify studies investigating prosody perception and production in individuals with ASD to provide a more comprehensive and integrated understanding of the topic.

The results across the 14 included studies are highly heterogeneous, with some suggesting that autistic individuals perceive affective prosody similarly to neurotypical controls (Leung et al., 2023; Lyons et al., 2014; Ben-David et al., 2019), while others report apparent deficits (Maslennikova et al., 2022; Hoffman et al., 2023; Globerson et al., 2015; Chiew et al., 2017). These discrepancies are particularly evident when comparing findings from recent studies using multimodal or ecologically valid stimuli with earlier or more narrowly focused behavioral research.

Several studies suggest emotion-specific differences: for instance, Wang and Tsao (2015) found differences in recognizing happiness but not other emotions; Martzoukou et al. (2017) identified differences in perception between the ASD group and the control group limited to surprise. Interestingly, autistic individuals were, in some cases, more accurate in identifying certain emotions like sadness and fear. Schelinski and Von Kriegstein (2019) also found emotion-specific differences in prosodic perception between ASD and NT groups. However, the ASD group had more difficulty in recognizing the emotions of sadness and fear. Lui et al. (2023), one of the most recent studies, investigated the perception of affective prosody in individuals without an ASD diagnosis but who, according to the AQ, had certain autistic traits, and again, interesting differences emerged. Indeed, individuals who scored higher on the AQ were less able to perform the emotion recognition task than people with lower AQ scores.

Studies evaluating how neurotypical listeners perceived the prosody of autistic speakers further underscored this complexity. Gibson et al. (2023) and Hubbard et al. (2017) reported mixed findings on pitch modulation and emotion recognition accuracy, suggesting that the communicative mismatch may be bidirectional.

This review has some methodological limitations that must be acknowledged. First, the review was not preregistered, which may reduce its transparency and reproducibility. Second, it is possible that relevant studies were missed due to search string limitations or publication bias. Only studies published in English were included, which may have excluded potentially informative work in other languages. Third, no formal assessment of study quality or risk of bias was performed. This choice limits the ability to weigh the strength of the conclusions drawn from individual studies.

A central contribution of this review lies in identifying how methodological differences across studies contribute to inconsistent findings on prosodic perception and production in ASD. The diversity of tasks, ranging from emotion labeling (e.g., Leung et al., 2023; Wang & Tsao, 2015) to tone discrimination (Globerson et al., 2015; Schelinski & von Kriegstein, 2019) to prosodic production assessments (Gibson et al., 2023; Hubbard et al., 2017), clearly impacts the type of data collected and the conclusions drawn. Likewise, the heterogeneity of stimuli, which includes emotionally neutral sentences (Gebauer et al., 2014), low-pass filtered speech (Chiew et al., 2017), and isolated syllables (Maslennikova et al., 2022), influences how emotional prosody is conveyed and perceived. For example, studies using ecologically valid or multimodal stimuli (e.g., Hoffmann et al., 2023) often report more pronounced difficulties in ASD groups compared to those employing simplified or artificial materials. In addition to variability in tasks and stimuli, the diversity of assessment instruments used across studies further complicates interpretation. Some studies employed standardized and structured tools (e.g., Lyons et al., 2014), enabling a more systematic comparison of prosodic skills across participants. Others relied on custom tasks or subjective ratings, as seen in Ben-David et al. (2019), where researchers used perceptual judgments without calibration. Neurophysiological methods like EEG (Maslennikova et al., 2022) and fMRI (Gebauer et al., 2014) offer more profound insights into processing mechanisms. However, they are less applicable in clinical or educational contexts due to their complexity and cost. Future research would benefit from integrating standardized behavioral assessments, ecologically valid stimuli, and objective acoustic and perceptual measures. Instruments that assess both production and perception may be particularly valuable in capturing the multifaceted nature of prosodic processing in ASD. Developing a core set of validated instruments could improve cross-study comparability and provide more precise clinical evaluation and intervention planning guidance.

These limitations suggest that a more unambiguous understanding of prosodic perception and production in ASD will require future studies using standardized research protocols. It is crucial to adopt more consistent and controlled methodologies to reduce the variability of

results. Furthermore, our findings have important implications for clinical practice and intervention strategies. Given the variability in prosodic abilities among autistic individuals, clinicians should employ individualized approaches when assessing and addressing prosodic difficulties. Interventions should target perceptual and production-related challenges, incorporating methods that enhance expressive and receptive prosodic skills. The inconsistency across studies also underscores the need for validated assessment tools to measure prosodic features reliably. Future research should prioritize the development of such tools and explore longitudinal changes in prosodic abilities to determine how targeted interventions influence prosodic perception and production over time.

Although this review has limitations, it highlights that the perception and production of prosody in individuals with ASD is still an open topic to explore. These findings emphasize the need for integrative, multimethod approaches to capture the prosodic profile of ASD individuals fully.

Conclusions

In conclusion, the conflicting findings in the reviewed studies underscore the complexity of prosodic perception and production in individuals with ASD. Theoretically, these results challenge existing models of emotional processing in ASD, suggesting that multiple factors, such as age, task design, and stimulus complexity, may influence the ability to perceive and produce affective prosody. These findings highlight the need for tailored interventions that account for the variability in prosodic abilities among autistic individuals. For instance, speech and language therapies could incorporate emotion-specific training to enhance prosodic perception and expression in social communication. Additionally, since neurotypical listeners may struggle to interpret the prosody of autistic speakers, increasing awareness and fostering communication strategies that accommodate these differences could improve social interactions. Future research should adopt standardized methodologies to clarify inconsistencies and better understand the underlying mechanisms of prosodic processing in ASD. If these findings are valid and replicable, they could contribute to more effective diagnostic assessments and intervention programs, ultimately improving the quality of life for individuals with ASD by enhancing their communicative and emotional expressivity.

Acknowledgements, Grants and Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest statement

Declarations of interest: none.

Authors' contribution

All authors contributed to and have approved the final manuscript.

References

- Ben-David, B. M., Ben-Itzhak, E., Zukerman, G., Yahav, G., & Icht, M. (2020). The perception of emotions in spoken language in undergraduates with high functioning autism spectrum disorder: A preserved social skill. *Journal of Autism and Developmental Disorders*, 50, 741-756. <https://doi.org/10.1007/s10803-019-04297-2>
- Bone, D., Black, M. P., Ramakrishna, A., Grossman, R. B., & Narayanan, S. S. (2015). Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism. In *Interspeech*, 1616-1620. DOI: 10.21437/Interspeech.2015-37
- Chiew, J., Kjelgaard, M., Chiew, J., & Kjelgaard, M. (2017). The perception of affective prosody in children with autism spectrum disorders and typical peers. *Clinical Archives of Communication Disorders*, 2(2), 128-141. <https://doi.org/10.21849/cacd.2017.00157>
- Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*, 74(6), 698. <https://doi.org/10.1037/amp0000399>
- Cutler, A., & Clifton, C. (1984). The use of prosodic information in word recognition. In H., Bouma, & D. G., Bouwhuis (Eds.), *Attention and performance X: Control of language processes*, 183-196. London: Erlbaum.
- Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, 27(3), 867-884. <https://doi.org/10.1017/S0954579414000741>
- Fay, W. H., & Schuler, A. L. (1980). *Emerging language in autistic children*. Baltimore: University Park Press.
- Fox Tree, J. E., & Meijer, P. J. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research*, 63(1), 1-13. <https://doi.org/10.1007/PL00008163>
- Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, 10(3), 384-407. <https://doi.org/10.1002/aur.1678>
- Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical*, 6, 370-378. <https://doi.org/10.1016/j.nicl.2014.08.025>
- Gibson, M. T., Schmidt-Kassow, M., & Paulmann, S. (2023). How neurotypical listeners recognize emotions expressed through vocal cues by speakers with high-functioning autism. *PLOS ONE*, 18(10), e0293233. <https://doi.org/10.1371/journal.pone.0293233>
- 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.
- Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. (2007). The 'Reading the Mind in the Voice' test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, 37, 1096-1106. <https://doi.org/10.1007/s10803-006-0252-5>
- Grossman, R. B., Bemis, R. H., Skwerer, D. P., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech,*

- Language and Hearing Research*, 53(3), 778-793. [https://doi.org/10.1044/1092-4388\(2009/08-0127\)](https://doi.org/10.1044/1092-4388(2009/08-0127))
- Grossman, R. B., & Tager-Flusberg, H. (2012). "Who said that?" Matching of low-and high-intensity emotional prosody to facial expressions by adolescents with ASD. *Journal of Autism and Developmental Disorders*, 42, 2546-2557. <https://doi.org/10.1007/s10803-012-1511-2>
- Hoffmann, J., Travers-Podmaniczky, G., Pelzl, M. A., Brück, C., Jacob, H., Hölz, L., ... & Wildgruber, D. (2023). Impairments in recognition of emotional facial expressions, affective prosody, and multisensory facilitation of response time in high-functioning autism. *Frontiers in Psychiatry*, 14, 1151665. <https://doi.org/10.3389/fpsy.2023.1151665>
- Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research*, 10(12), 1991-2001. <https://doi.org/10.1002/aur.1847>
- Hübscher, I., & Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development. *Frontiers in Psychology*, 10, 1259. <https://doi.org/10.3389/fpsyg.2019.01259>
- Leung, F. Y., Stojanovik, V., Micai, M., Jiang, C., & Liu, F. (2023). Emotion recognition in autism spectrum disorder across age groups: A cross-sectional investigation of various visual and auditory communicative domains. *Autism Research*, 16(4), 783-801. <https://doi.org/10.1002/aur.2896>
- Lui, M., Lau, G. K. B., Han, Y. M. Y., Yuen, K. C. P., & Sommer, W. (2023). Strong relationship between rapid auditory processing and affective prosody recognition among adults with high autistic traits. *Journal of Autism and Developmental Disorders*, 53(8), 3180-3193. <https://doi.org/10.1007/s10803-022-05600-4>
- Lyons, M., Schoen Simmons, E., & Paul, R. (2014). Prosodic development in middle childhood and adolescence in high-functioning autism. *Autism Research*, 7(2), 181- 196. <https://doi.org/10.1002/aur.1355>
- Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of autism and developmental disorders*, 34, 311-328. <https://doi.org/10.1023/B:JADD.0000029553.52889.15>
- Martoukou, 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, 1573-1595. <https://doi.org/10.1007/s10936-017-9500-4>
- Maslennikova, A. V., Portnova, G. V., & Martynova, O. V. (2022). Brain oscillatory patterns of affective prosody perception in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 96, 101993. <https://doi.org/10.1016/j.rasd.2022.101993>
- Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 35(3), 204-214. [https://doi.org/10.1002/\(SICI\)1098-2302\(199911\)35:3<204::AID-DEV5>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2302(199911)35:3<204::AID-DEV5>3.0.CO;2-V)
- McDonald, S., Togher, L., Tate, R., Randall, R., English, T., & Gowland, A. (2013). A randomised controlled trial evaluating a brief intervention for deficits in recognising emotional prosody following severe ABI. *Neuropsychological rehabilitation*, 23(2), 267-286. <https://doi.org/10.1080/09602011.2012.751340>
- Mesibov, G. B. (1992). Treatment issues with high-functioning adolescents and adults with autism. In E. Schopler & G. B. Mesibov (Eds.), *High-Functioning individuals with autism*, 143–155. Springer. https://doi.org/10.1007/978-1-4899-2456-8_8
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Prisma-P

- Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, 4, 1-9. <https://doi.org/10.1186/2046-4053-4-1>
- Patel, S. P., Nayar, K., Martin, G. E., Franich, K., Crawford, S., Diehl, J. J., & Losh, M. (2020). An acoustic characterization of prosodic differences in autism spectrum disorder and first-degree relatives. *Journal of Autism and Developmental Disorders*, 50, 3032-3045. <https://doi.org/10.1007/s10803-020-04392-9>
- Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of autism and developmental disorders*, 35, 205-220. <https://doi.org/10.1007/s10803-004-1999-1>
- Peppé, S., Cleland, J., Gibbon, F., O'Hare, A., & Castilla, P. M. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics*, 24(1), 41-53. <https://doi.org/10.1016/j.jneuroling.2010.07.005>
- Pfanner, L., Tancredi, R., & Marcheschi, M. (2008). Comunicazione e linguaggio nei disturbi pervasivi dello sviluppo. *Giornale di Neuropsichiatria dell'Età Evolutiva*, 28, 59-74.
- Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. D. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 39(2), 221-257. <https://doi.org/10.1017/S030500091100002X>
- 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, 68-82. <https://doi.org/10.1007/s10803-018-3681-z>
- Teresa Bruno, M., Russo, S., & Valenza, E. (2022). Abilità prosodiche e competenze sociopragmatiche nel Disturbo dello Spettro Autistico. Una rassegna sul ruolo delle atipicità attentive. *Psicologia clinica dello sviluppo*, 26(2), 227-256. DOI: 10.1449/103812
- Wang, J. E., & Tsao, F. M. (2015). Emotional prosody perception and its association with pragmatic language in school-aged children with high-function autism. *Research in developmental disabilities*, 37, 162-170. <https://doi.org/10.1016/j.ridd.2014.11.013>
- Yankowitz, L. D., Schultz, R. T., & Parish-Morris, J. (2019). Pre-and paralinguistic vocal production in ASD: Birth through school age. *Current psychiatry reports*, 21, 1-22. <https://doi.org/10.1007/s11920-019-1113-1>