Systematic Review of Linguistic Characteristics in Profiling and Automated Detection of Autistic Speech

Charlotte Bellinghausen[1], Andreas Riedel[2,3]
[1]Institute of German Studies, Linguistics, University of Duisburg-Essen, Germany
[2]Department of Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany
[3]Luzerner Psychiatrie, Ambulante Dienste, Löwengraben 20, 6004 Luzern, Switzerland
charlotte.bellinghausen@uni-due.de, andreas.riedel@lups.ch

Abstract

Our work aims to provide a systematic review of interdisciplinary studies on the automated recognition and differentiation of autistic and non-autistic speech. Linguistic patterns of autistic speech can serve as indicators for clinical diagnosis. Recognizing and integrating autistic language patterns can help to develop systems characterized by neurodiversity in language use. We selected 24 studies and systematically compared them to find similarities and differences between autistic and non-autistic groups, focusing on linguistic analysis. We divided the studies into three types: automatic recognition studies, language profile studies and meta-analysis studies. The vast majority of studies investigated the acoustic modality for English. The material mostly consisted of children's speech, but the meta-analyses included different age groups. Linguistic analysis indicated that, alongside semantic, pragmatic, and lexical features, prosodic and acoustic cues were especially salient in the various speech categorization approaches. These findings highlight prosodic and acoustic features as central to the identification of autistic speech.

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by severe impairments in the areas of social interaction and social communication, as well as by restricted, repetitive, and inflexible behaviours and interests or activities (ICD-11 'autism', 2025). In recent years, public awareness and early identification account have contributed to increasing prevalence rates over time (cf. Zeidan et al., 2022: 8). The World Health Organization (WHO, 2025) estimates that the global prevalence of autism is approximately 1 in 100 children. In epidemiological samples, the male-to-female ratio is approximately 3:1. However, there are concerns about under-reporting in girls and women (cf. American Psychiatric Association 2022: 64).

While the cause of autism is largely genetic, other factors may also play a role (cf. Kamp-Becker & Bölte, 2014: 33ff). Markowetz (2020: 14) gives an overview of approaches and theories to explain the causes of autism, referring to neurobiological and psychosocial reasons.

According to ICD-11 and DSM-5, autism is no longer divided into subcategories, the subtypes of autism are subsumed under the umbrella term of *autism spectrum disorder*. Following a neurodiversity perspective, we use the terms *autism spectrum* and *autism* in the following.

The *autism spectrum* can be characterised by specific forms of expression, peculiarities, and abilities (cf. Theunissen, 2020: 55). According to Eberhardt-Juchem (2023: 29), the range of linguistic and communicative abilities in people of the autism spectrum is wide. For example, there may be functional verbal language with or without impaired language development. In other cases, there are no or minimal verbal skills (cf. Eberhardt-Juchem, 2023: 30). For a review of language in autism, see Lindmeier et al. (2024) and Riedel and Biscaldi-Schäfer (2025, in press).

In cases of autism spectrum without intellectual impairment, the processing on the syntactic and semantic level is barely affected, but problems in pragmatic interpretation are often found (cf. Riedel, 2015: 4f., see also Bellinghausen et al., 2024: 2).

Proceedings of the 21st Conference on Natural Language Processing (KONVENS), Volume 1: Long and Short Papers, pages 269–281, 2025

As Bellinghausen et al. (2024: 2) pointed out, problems in general pragmatic processing are often found in autism without intellectual impairment. However, there is empirical evidence that the pragmatic abilities of hearers with autism spectrum without intellectual impairment differ between pragmatic domains (see also Geurts et al., 2019: 114 and Andrés-Roqueta & Katsos, 2020).

The role of prosody in autistic persons without intellectual impairment has been investigated in several studies, covering both speech production and perception (for a description see Bellinghausen et al., 2024: 2-3).[1] Some studies suggest evidence for a difference in both prosody production (e.g. Shriberg, 2001; Kiss, 2012) and perception (Diehl & Paul, 2012). The findings differ with respect to the type of prosody. For prosodic uncertainty perception no significant difference between the autistic and neurotypical group was found (see Bellinghausen et al., 2024). Globerson (2015) reported no group differences in structural prosody perception, but in emotional prosody perception of the prototypical emotions happiness, sadness, anger, and fear. Due to the different functions of prosody, differences between autistic and neurotypical groups may vary between studies (cf. Bellinghausen et al., 2024: 3).

This raises the question of how differences in speech production can be used for computer-assisted diagnosis. Another question is how to recognise autistic speech automatically. In their survey of autism detection in speech, Neitzel et al. (2018) provided a description of the LENA system for the analysis of autistic speech in current research. It can be used to assess the number of words spoken by adult speakers, the number of vocal utterances spoken by children, and the number of speaker changes (cf. Neitzel et al., 2018: 86, see Ganek & Eriks-Brophy, 2018 for an overview of the use of the LENA system). AI-based systems have the potential to become valuable tools for clinical diagnosis in medicine, psychology and speech therapy.

The goal of the current study is to provide a systematic review of interdisciplinary studies on the characteristics of autistic and neurotypical speech, focusing on linguistic analysis.[2] Linguistic patterns of autistic speech can serve as indicators for clinical diagnosis. The automated recognition of these patterns can be helpful for intervention. By integrating perspectives from linguistics, computer science, psychology, and medicine, our primary aim is to identify the linguistic features that are characteristic of autism, while also highlighting key similarities, differences, and research gaps. Our comparison of studies will focus on the linguistic levels considered in autistic and neurotypical speech analysis and detection. Our findings aim to advance speech technology and natural language technology by bridging these disciplines. An overview of the different disciplines covered by our work is given in Table 1.

discipline	**subject for the current project**
linguistics	description of language at various levels: syntax, semantics, pragmatics, lexis, phonetics (prosody), etc.
medicine & psychology	autism research as a broad topic; studying similarities between autistic and neurotypical groups
speech technology & natural language technology	provision of a range of tools for the analysis of spoken and written language; use of artificial intelligence tools

Table 1: Role of different disciplines in the current paper

Integrating AI-based automated detection of autistic speech with human intervention may yield improved outcomes.

2 Method

2.1 Paper selection

We selected 24 interdisciplinary studies from the fields of linguistics, computer science, psychology, and psychiatry. The resulting studies were published between 2011 and 2024. Studies were mainly retrieved from the ISCA (International Speech Communication Association) archive (ISCA, 2025), from the PubMed database (PubMed, 2025), and from

[1] For an overview of the functions of both structural and emotional prosody, see Grice et al. (2023).

[2] The preliminary results of this work were presented at the annual conference ‘Wissenschaftliche Tagung Autismus Spektrum 2025’ with a published abstract (Bellinghausen et al., 2025).

Google searches. Keyword searches included different combinations of ‘autism’, ‘automatic classification/detection/discrimination’, ‘speech’, ‘language profile’, ‘prosody’, and others.

Our review includes studies that assessed speech using pre-defined features as well as studies that used AI tools for automatic recognition.

2.2 Evaluation criteria

To obtain a systematic overview of the 24 selected studies, we first categorise them according to study type in sec. 3.1. Sec. 3.2 forms the core of our analysis. Under the heading 'Linguistic Analysis', we analyse the different linguistic levels, features, and methods employed in the selected studies, including semantics, pragmatics, lexis and prosody. Table 1 (see Appendix) provides information about the participants, on the tested languages, and the linguistic analysis for each selected study.

3 Results of Systematic Evaluation

3.1 Study characteristics: type of study

The different study types included in our analysis are shown in the diagram in figure 1. The majority of studies were assigned to the category 'automated detection study for the acoustic modality' (N=12; Ashwini et al., 2023; Xu et al., 2023; Yang et al., 2023; MacFarlane et al., 2022; Lee et al., 2013; Chi et al., 2022; Asgari et al., 2021; Prud'hommeaux & Rouhizadeh, 2012; Xu et al., 2009; Lahiri et al., 2023; Asgari et al., 2013). We found one study on ‘automated detection’ for the visual modality (N=1; Li et al., 2022).

Studies of language profiles occurred with N=6 (Geurts & Embrecht, 2008; Volden et al., 2009; Philofsky et al., 2007; Eni et al., 2023; Black et al., 2011; Wetherby et al., 2007).

There are N=5 studies in the category ‘overview/review studies/meta-analysis’ (Fusaroli et al, 2017; Asghari et al., 2021; Probol & Mieskes, 2024; Ma et al., 2024; Vogindroukas et al., 2022).

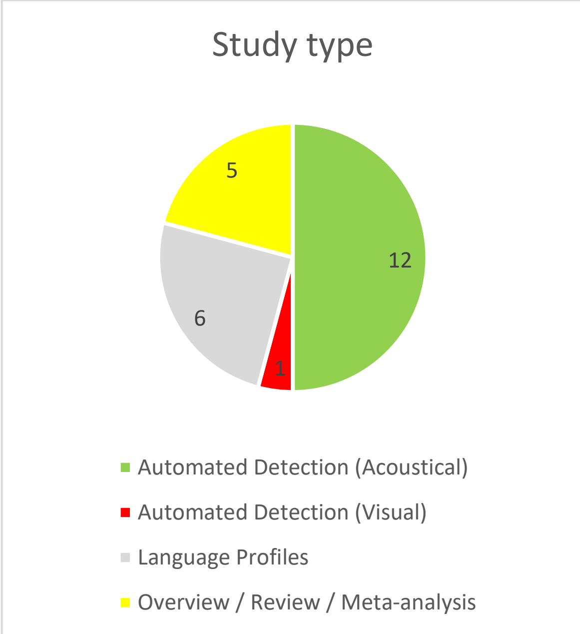

Figure 1: Different study types, N=24

3.2 Linguistic analysis

The results of the systematic analysis of the automated detection studies are described in sec. 3.2.1. The analysis of profile studies is presented in sec. 3.2.2, the analysis of systematic review studies follows in sec. 3.2.3. Other studies are discussed in sec. 3.2.4.

3.2.1 Studies on automatic detection

The results of our systematic analysis of studies on automated speech detection are shown in Table 2 (see Appendix). The table contains the following criteria: i) study reference, 2) participants, 3) language tested, 4) linguistic level/method, 4) linguistic features, 5) results. The first group of studies are those that compare the language of an autistic group with that of a neurotypical comparison group. We first refer to studies in the **acoustic modality**.

Ashwini et al. (2023) used semantic-pragmatic features for automatic detection for the acoustic modality. Selected features included echolalia, coherence, and repetitive speech. The results suggested that semantic and pragmatic language features can serve as key features for autism diagnosis complementing syntactic and lexical features, and achieving 94% classification accuracy.

Both Cho et al. (2022) and Yang et al. (2023) tested acoustic and lexical features for the automatic detection of speech from autistic and neurotypical individuals. Cho et al. (2022) used a gradient boosting model and principal component analysis for automatic classification. The features used were frequency-based, voice quality, energy and loudness measurements, and statistical features. The final model's accuracy was above chance levels, reaching 75%.

Yang et al. (2023) selected pitch- and voice-quality-related features, as well as the percentage of silences and lexical features such as ratio measurements, to develop various classifiers. The most effective classifier was a support vector machine with an accuracy rate of 61%, surpassing the performance of human raters (49%). Furthermore, a pre-trained audio transformer was used to extract acoustic features, achieving a prediction rate of 68%.

MacFarlane et al. (2022) used automated language and voice measures in their studies. By analysing language and voice features from ADOS-2 transcripts and audio recordings using automated language and voice measures (ALMs and AVMs), the study successfully classified speech from autistic and non-autistic children. The AVM model achieved an area under the curve (AUC)-ROC of 0.7800, the ALM model achieved an AUC-ROC of 0.8748, and the combined model demonstrated a significant improvement with an AUC-ROC of 0.9205.

Lee et al. (2013) used the following ensemble classification techniques: support vector machines (SVMs), deep neural networks (DNNs), k-nearest neighbors (KNNs) and the acoustic segment mode (ASM) technique. Acoustic features were extracted using the open SMILE feature extractor. The results showed that the final ensemble system, which combined machine learning and ASM techniques, outperformed the challenge baseline.

Chi et al. (2022) used machine learning techniques to classify autistic and non-autistic speech using speech recognition features. The random forest classifier achieved 70% accuracy; the fine-tuned Wav2vec 2.0 model, 77%; and the convolutional neural network, 79%.

Asgari et al. (2021) employed speech recognition features to distinguish between autistic and neurotypical speech. Standard speech measures included loudness, cepstral coefficients, and spectral entropy. The results of these models suggested that they performed better than a chance model. However, articulation measures alone achieved only moderate classification performance, with an AUC-ROC of around 68%. By contrast, prosodic features (e.g. pitch, intonation and rhythm) performed significantly better, achieving an AUC-ROC of around 82% which increased to 83% when the voice recording context was optimized**.**

Prud'hommeaux et al. (2012) used pragmatic features related to relevance and topicality to categorise speech into different narratives. Using automatically derived features, they distinguished between autistic and non-autistic speech. However, they did not investigate the correlation with human judgements of the features. They reported accuracy as the area under the receiver operating characteristic curve.

Xu et al. (2009) used pattern recognition and machine learning techniques on data from children with and without autism, as well as those with language delays. Autism was detected with an accuracy rate of between 85% and 90% in cross-validation tests.

Having referred to the studies on the acoustic modality, we will now turn to the **visual modality**. The study by Li et al. (2022) focused on visual recognition using features such as facial expression, head pose, and head trajectory features. The accuracy of automatic recognition was 96%.

The second group of detection studies includes the work of Xu et al. (2023) and Lahiri et al (2023). Xu et al. (2023) performed binary classifications for the audio condition. They tested whether i) autistic child speech could be distinguished from neurotypical adult speech and whether ii) speech or non-vocalisations occurred. Linguistic levels and features were not explicitly described. The results showed F1 macroscores of 83% for i) and 68% for ii). Lahiri et al. (2023) presented work on the automatic classification of child-adult speakers. The children were autistic speakers and the adults were neurotypical speakers. The model was pre-trained using child-inclusive interactions, based on two recent self-supervision algorithms: Wav2vec 2.0 and WavLM. The authors achieved a relative improvement of 9–13% over the state-of-the-art baseline in F1 classification scores on two clinical interaction datasets involving autistic children. However, it should be noted that a detailed description of the linguistic levels and features was lacking.

The third subtype of classification studies subsumes only the study by Asgari et al. (2013). In this study, the speech of autistic and non-autistic children was detected acoustically. Four subtypes of autism were classified in the sample of autistic children using voice quality-related features derived from harmonic analysis. The features improved performance by 2.3% in detecting autism spectrum disorder and by 2.8% in classifying it into four categories compared to the baseline results, measured by unweighted average recall (UAR).

3.2.2 Profile studies

Table 3 (see Appendix) contains the following information: 1) study reference, 2) language, 3) method, 4) linguistic levels/features, and 5) results. The table consists of two groups: studies for i) the audio and ii) the audiovisual modality. i) Geurts et al. (2008) used the Children's Communication Checklist to assess structural and pragmatic language in children with autism, SLI, and ADHD. School-aged children with autism and ADHD showed similar profiles with more pragmatic than structural difficulties. In contrast, preschool children with autism and SLI had more structural language problems.

Using multiple regression analyses, Volden et al. (2009) investigated the relationship between non-verbal cognition, structural language, and pragmatic language in autistic children. They found that structural language skills predicted pragmatic performance; however, they also found that variance in pragmatic scores could not be accounted for by structural language or non-verbal cognition.

In the study by Philofsky et al. (2007), parents rated the language skills of their children with autism or Williams syndrome using the Children's Communication Checklist, a standardised assessment tool for pragmatic language, including both verbal and non-verbal communication. Problems with pragmatic language functioning were observed in both groups, but children with Williams syndrome performed significantly better on the coherence, stereotyped language, non-verbal communication and social relations subscales.

Eni et al. (2023) estimated autism severity using hand-crafted features and log-mel spectrograms. In most cases, the hand-crafted features produced lower prediction errors than the log-mel spectrograms. The best results were achieved when both feature types were combined.

(ii) Language profiles in the audiovisual modality have been investigated by Black et al. (2011) and Whetherby et al. (2007). Black et al. (2011) examined the social communication profiles of children with and without autism and with developmental delay, as observed in videotaped behavioural samples. In their initial analysis, they annotated the speech with respect to lexical and phonetic information, as well as non-verbal communication, vocalisations, and sentence types. This paper does not describe the results.

Whetherby et al. (2007) used multimodal video samples to compare the social communication of autistic, neurotypical and language-delayed children. Using 14 Communication and Symbolic Behavior Scales (CSBS) measures (e.g. gaze shift and shared positive affect), they found significant differences between autistic and neurotypical children, but fewer differences between autistic and language-delayed children.

3.2.3 Review studies and meta-analysis studies

The next step is to analyse the review studies and meta-analysis studies (N=5, see Table 3). Table 3 contains the following information: 1) study reference, 2) language, 3) method, 4) linguistic levels/features, and 5) results.

Both Fusaroli et al. (2017) and Asghari et al. (2021) conducted systematic reviews and meta-analyses of studies investigating acoustic and prosodic patterns in autistic and neurotypical groups. Fusaroli et al.'s (2017) included 34 studies in their work, comprising 30 univariate studies aiming at identifying differences between the two groups, and 15 multivariate machine-learning studies. Their results for the univariate studies revealed group differences in mean pitch and pitch range; these studies examined one acoustic feature at a time. The results of the multivariate showed a higher accuracy rate by 63-96% compared to the univariate studies.

By contrast, Asghari et al. (2021) systematically compared 39 studies that tested autistic participants from different age groups. They found that autistic participants exhibited a higher mean pitch, greater pitch range, longer voice duration and greater pitch variability than the neurotypical comparison group.

Probol and Mieske's (2024) work is a survey and meta-analysis. They presented studies that examined verbal and semantic fluency, prosodic features, and also disfluencies and speaking rate. The results suggested an under-representation of females. Most of the papers focused on traditional learning approaches rather than transformers and deep learning techniques.

Ma et al. (2024) systematically compared the results of empirical studies on acoustic analysis and machine learning techniques. The aim was to apply natural speech prosody to the detection of autistic speech based on statistically supported evidence. The prosodic analysis included the features pitch mean, pitch range, pitch standard deviation, pitch variability, utterance duration, speaking rate, intensity mean and variation. The authors found significant differences in the pitch-related measures, but not in the temporal features. Multivariate machine learning trained on natural speech samples showed accuracy in detecting ASD with a sensitivity of 76% and a specificity of 80%.

Vogindroukas et al. (2022) aimed to define the language profiles of autistic individuals. The linguistic levels considered were semantics, pragmatics, phonology, morphosyntax covering spoken and written language, as well as non-verbal behaviour. The analysis resulted in four subtypes of profiles depending on language impairment, comorbidity, social communication and social interaction. It was concluded that problems with the semantic aspect of language influenced the skills in the areas of abstract thinking, word ambiguity, concept categorization, etc.

4 Discussion

We presented a systematic analysis of 24 selected interdisciplinary studies on the recognition and differentiation of autistic and neurotypical speech. Taking a linguistic perspective, we analysed the underlying linguistic features used to characterise autistic speech, identifying similarities and differences between autistic and non-autistic speech. Firstly, we divided the studies into three categories: 1) studies on automatic detection; 2) language profile studies; and 3) meta-analyses and systematic reviews. Most studies investigated speech samples from children and adolescents, whereas the meta-analyses included all age groups. English was the language tested in most studies.

Our summary of the selected studies for the three subtypes is as follows: 1) **Studies on automated detection:** The vast majority of studies investigated the audio modality. Automatic detection encompassed the distinction between autistic and non-autistic groups, child and adult speech, and autistic subtypes. Linguistic levels included the lexical, semantic, and pragmatic levels, as well as prosodic and other acoustic cues. Classification accuracy ranged from 61% to 94%, depending on the combination of linguistic cues used. Prosodic features, such as pitch**,** intonation**,** rhythm**,** and speech rate, play a critical role in distinguishing autistic from neurotypical speech (Asgari et al., 2021; Yang et al., 2023). Combining semantic-pragmatic features (e.g. coherence, repetition, echolalia and topic relevance) yielded an accuracy rate of 94% (Ashwini et al., 2023). Similarly, combining acoustic and lexical features (e.g. voice quality, energy and loudness measurements, frequency-based features, and statistical features) achieved an accuracy rate of 75% (Cho et al., 2022). Asgari et al. (2021) achieved an AUC-ROC of 83%, which was significantly higher than that achieved using purely articulatory features (e.g. loudness, cepstral coefficients, spectral entropy). Visual features (e.g., facial expressions) can also achieve very high accuracy (up to 96%). In terms of automatic classification methods, support vector machines, random forests, gradient boosting, convolutional neural networks (CNNs) and deep neural networks were employed. In the work of Chi et al. (2022), CNN achieved 79% accuracy and Wav2Wav 2.0 reached 77%.

In summary, it can be concluded that prosodic and pragmatic features are effective in discriminating between autistic and neurotypical speech. Multimodal and hybrid feature sets (acoustic, lexical, pragmatic and visual) yielded high classification rates. However, it should be noted that in some studies, it was not described which linguistic features were used. One possible explanation for this is that current AI systems do not use predefined linguistic features for recognition, but rather learn the characteristics on which they base their decisions from the training data. In order to examine what features those systems had extracted for classification, the framework of explainable artificial intelligence (XAI) can be used.

2) **Profile studies:** The features used for profiling were mostly based on standardised checklists, or were manually annotated. These covered both acoustic and visual modalities. However, the selected studies do not clearly and consistently demonstrate the differences and

similarities between autistic and non-autistic groups. However, there is evidence that pragmatic language impairments are prevalent among school-aged children with autism. Studies by Geurts et al. (2008) and Philofsky et al. (2007) reported difficulties with pragmatic language skills in autistic children. Combining hand-crafted acoustic-prosodic features with log-mel spectrograms improved the accuracy with which autism severity could be estimated (Eni et al., 2023).

3) **Meta-analysis studies:** It should be emphasised that the individual factors at the methodological and linguistic levels depend on the studies selected for each review. A key finding is that pitch-related prosodic features (e.g. higher mean pitch, greater pitch range, and pitch variability) are consistently altered in autistic speech (Fusaroli et al., 2017; Ashgari et al., 2021; Ma et al., 2024). Furthermore, multivariate machine learning approaches using speech prosody outperformed univariate methods (Fusaroli et al., 2017; Ma et al., 2024). Another finding is that female autism profiles are under-explored (Probol & Mieskes, 2024).

In summary, we found that most, but not all studies considered pragmatic, lexical, prosodic and other acoustic cues. Further work should systematically include linguistic levels in the analysis of autistic and non-autistic speech in order to provide a more detailed picture of the of autistic speech characteristics. Such research could contribute to a better understanding of the role and interplay of different linguistic features in automatic classification.

In future studies, it would be very interesting to further investigate the different modalities in a larger number of selected studies, as well as using speech material from different languages and multilingual corpora.

We also recognise the importance of AI-based speech systems as valuable tools for clinical diagnosis in medicine, psychology, and speech therapy. By highlighting their potential, we aim to advocate for their broader adoption and further development. For more information on the role of artificial intelligence sound classifiers in autistic speech, see also Chi et al. (2022: 2).

In our view, speech technology tools can provide valuable guidance and support for the experienced human clinical diagnosticians, but they cannot replace them. Future work could involve comparing the autistic group with other psychiatric groups. In this way it might be possible to identify similarities and differences between the groups and use the findings for differential diagnosis.

5 Acknowledgement

We would like to thank Bernhard Schröder, Lisa Korte, Ulrich Schade, and Nantke Pecht for their helpful comments on this paper.

6 Appendix

classification type / study	**speakers**	**language**	**linguistic level / *method***	**linguistic features**	**results**
Ashwini et al. (2023)	AS: N= 35; NTC: N=41; 3-8 y.	English	semantic-pragmatic / *NLP and machine learning techniques*	echolalia, semantic coherence, repetitive language	94% accuracy
Cho et al. (2022)	AS: N=35; NTC: N=25, average age: 11 y.	English	acoustic / lexical / *gradient boosting model, principal component analysis*	shimmer, Harmonic-to-noise ratio, energy, Mel Frequency Cepstral Coefficients / word frequencies, pronoun usages, number of filler words	75% accuracy
Yang et al. (2023)	AS: N=35; NTC: N=37; average age: 11; 10 y.	English	acoustic / lexical / *machine learning approach*	e.g. percent of silence, word and sentence ratio	support vector machine: 61% accuracy, transformer: 68%
MacFarlane et al. (2022)	AS: N=88; NTC: N=70; 7-17 y.	English	*NLP methods, harmonic model of speech*	Automated voice measures (AVM) and automated language measures (ALM)	AVM model AUC-ROC: 0.7800; ALM model an AUC-ROC: 0.8748; combined model: 0.9205
Lee et al. (2013)	not specified	not specified	*ensemble classification techniques: support vector machines, deep neural networks, k-nearest neighbours; acoustic segment mode technique (ASM)*	acoustic features extracted by open SMILE feature extractor	final ensemble system combination (machine learning and ASM techniques) outperformed the challenge baseline
Chi et al. (2022)	AS: N=20, NTC: N=38, 3-12 y.	English	*machine learning techniques: random forests, convolutional neural networks (CNNs), fine-tuned wav2vec 2.0 models*	speech recognition features	random forest classifier: 70% accuracy, fine-tuned wav2vec 2.0 model: 77% accuracy, convolutional neural network: 79% accuracy
Asgari et al. (2021)	AS: N=90; NTC: N=28; average age: 11 y.	English	*harmonic model of voiced signal*	standard speech measures: loudness, cepstral coefficients, spectral entropy	articulation measures alone: AUC-ROC: 68%; prosodic measures: AUC-ROC: 82-83%
Prud'hommeaux & Rouhizadeh (2012)	SLI N=17; AS*: N=21; NTC: N=40; 4-8 y.	English	pragmatic / *support vector machine*	relevance and topicality	distinction between autistic and non-autistic speech; accuracy as AUC-ROC
Xu et al. (2009)	AS: N=34; LD: N=30; NTC: N=76; age: not specified	English	acoustic / *machine learning, pattern recognition*	high dimensional features	accuracy of 85% to 90% in cross-validation tests for AS vs. LD and ASS vs. LD and NCT
Li et al. (2022)	AS: N=136; NTC: N=136 4-7 y.	not specified	non-verbal / novel attention-based facial expression recognition algo-rithm	facial expression, head pose, head trajectory	96% accuracy
Xu et al. (2023)	45 sessions (child/parent): average age: 7 y.	English	annotation categories: 1) child vs. adult speech, 2) intelligible, unintelligible, vocalization / *classification*	speech and nonverbal vocalizations	F1 macro scores: 83% and 68%
Lahiri et al. (2023)	346 sessions from AS: N=86; NTC: N=79; 3-13 y.	not specified	*pretraining of data with self-supervision algorithms: Wav2vec 2.0 and WavLM*	not specified	9-13% relative improvement over the state-of-the-art baseline
Asgari et al. (2013)	99 children; 9-18 y.	French	acoustic / *harmonic noise model*	voice quality, energy, spectral features, cepstral features	improvement in terms of unweighted average recall

Table 2: systematic overview of studies on automatic detection, N=13, NLP=natural language processing, AS: autism spectrum group: NTC: neurotypical comparison group; LD: language delay group, SLI: specific language impairment group; AS: autism spectrum group without language impairment; y=years, AUC-ROC: area under the receiver operating characteristics curve, AVM: automated voice measures, ALM: automated language measures

study: topic	speakers	language	Method	linguistic levels / features	results
Geurts et al. (2008): generation of language profiles in AS, SLI, and ADHD	ADHD: N=29; AS: N=29; study 1: 7-14y.; study 2: 4-7 y.	Dutch	Children's Communication Checklist	mainly pragmatics: use of context, non-verbal communication	AS & ADHD: similar language profiles
Volden et al. (2009): factors related to pragmatic development in AS	37 children: 6–13 years;	English	regression analysis	non-verbal cognition, structural language and pragmatic language	structural language skills predicted pragmatic performance
Philofsky et al. (2007)	AS: N=22; WS: N=21, NTC: N=19; average age: 6-9 y.	English	Children's Communication Checklist	verbal and non-verbal communication	better pragmatic functioning in WS
Eni et al. (2023): autism severity estimation by means of hand-crafted features and log-mal spectogram	AS: N=77%, NTC: N=23% N=127; 4.09±1.31 y.	Hebrew	feature extraction methods: implemented and tested for estimating AS severity score; five system configurations: tested for AS severity estimation	acoustic-prosodic features: pitch, formants, bandwiths, voicing, energy, spectral slope, ZCR, jitter, duration, quantity	1) hand-crafted features: lower prediction error in most configurations compared to log-mel spectrograms. 2) fusing autism severity scores for the two feature extraction methods yielded to best results
Black et al. (2011): presentation of multimodal corpus (USC CARE Corpus)	in total: N=60; 5-18 y.	English	audio-visual data collection during ADOS	lexical information; phonetical information (pronounciation), non-verbal communication and vocalizations; labeling as question, fragment, interruption, and/or complete thought	not described (work in progress)
Wetherby et al., (2007): social communication profiles from behavior samples videotaped	AS: N=50; DD: N=23, NTC: N= 50; age: 18-24 months	not specified	video recordings; parent/observer-child interaction	14 items of the CSBS Behavior Sample that form the social, speech, and symbolic composites: gaze shift, shared positive affect, gaze/point follow etc.	AS group: lower scores in 5 measures (vs. DD), lower scores in 14 measures (vs. NTC)

Table 3: Systematic analysis of profile studies, N=6; autism spectrum group; SLI: specific language impairment group, ADHD: attention deficit hyperactivity disorder; WS: williams syndrome; ZCR: Zero-Crossing Rate, DD: developmental delay, NTC: neurotypical comparison group

study: topic	**speakers & language**	**method**	**linguistic levels / features**	**results**
Audio				
Fusaroli et al. (2017): systematic review of literature: acoustic patterns in AS	study dependent	systematic review of 34 studies; childhood, adolescence, adulthood / *machine learning*	prosody: pitch, intensity, duration, speech rate, pauses, voice quality	univariate studies: AS vs. NTC: differences in mean pitch and pitch range, multivariate studies: 63-96%
Asghari et al. (2021): systematic review and meta-analysis AS vs. NTC: prosodic cues	study dependent	systematic review and meta-analysis of 39 studies; childhood, adolescence, adulthood	mean pitch, pitch range and variability, speech rate, intensity and voice duration	AS vs. NTC: differences in mean pitch, pitch range, voice duration and pitch variability
Probol & Miekes, (2024): survey and meta-analysis regarding linguistic, prosodic and acoustic cues as indicators of AS	study dependent	survey and meta-analysis of 10 studies; childhood, adolescence, adulthood	verbal and semantic fluency, prosodic features, but also disfluencies and speaking rate	1) underrepresentation of females, 2) most works: traditional machine learning approaches (SVMs, Naive Beyes, Linear Regression), few studies; transformer, deep learning
Ma et al. (2024): state-of-the-art review: analysis of prosodic features in spontaneous speech in AS and machine learning techniques	study dependent	state-of-the-art review of 25 eligible uni-variate studies on spontaneous speech and 18 multivariate studies on machine learning; childhood, adulthood	prosodic analysis: pitch Mean, pitch range, pitch standard Deviation (p.8), pitch variability, utterance duration, speaking rate, intensity mean and variation; machine learning techniques	1) AS vs. NTC: significant differences in pitch-related measures (but not in temporal features); 2) pitch range possibly influenced by age groups 3) machine learning trained on natural speech samples; achieved promising accuracy in detecting AS sensitivity: 75.51% specificity: 80.30%
Vogindroukas et al. (2022): language profiles in AS; try to classify these profiles according to combination of communication difficulties	study dependent	description of different studies	semantics, pragmatics, phonology, morphosyntax covering spoken, written language, non-verbal behavior	four subtypes of profiles depending on linguistic impairment, comorbidity, social communication, and social interaction

Table 4: systematic analysis of systematic review studies, N=5

References

American Psychiatric Association. 2022. Diagnostic and statistical manual of mental disorders. *Text Revision: DSM-5-TR*, Fifth Edition. American Psychiatric Association Publishing, Washington, DC.

Andrés-Roqueta, C., Katsos, N.A 2020.. A distinction between linguistic- and social-pragmatics helps the precise characterization of pragmatic challenges in children with Autism Spectrum Disorders and Developmental Language Disorder. *J Speech Lang Hear Res.,* 63:1494–508. doi: 10.1044/2020_JSLHR-19-00263

Asgari, M., Bayestehtashk, A., Shafran, I. 2013. Robust and accurate features for detecting and diagnosing autism spectrum disorders. In *Proceedings of Interspeech 2013*, Lyon, France, 191-194, doi: 10.21437/Interspeech.2013-65

Asgari, M., Chen, L., Fombonne, E. 2021. Quantifying Voice Characteristics for Detecting Autism. Front Psychol., 2021 Sep 7, 12:665096. doi: 10.3389/fpsyg.2021.665096

Asghari S.Z., Farashi, S., Bashirian, S., Jenabi, E. 2021. Distinctive prosodic features of people with autism spectrum disorder: a systematic review and meta-analysis study. *Sci Rep.*, 2021 Nov 29, 11(1):23093, doi: 10.1038/s41598-021-02487-6

Ashwini, B. Narayan, V., Shukla, J. 2023. SPASHT: Semantic and Pragmatic Speech Features for Automatic Assessment of Autism. In *Proceedings of ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing*, Rhodes Island, Greece, 1-5, doi: 10.1109/ICASSP49357.2023.10095135

Bellinghausen, C., Schröder, B., Rauh, R., Riedel, A., Dahmen, P., Birkholz, P., Tebartz van Elst, L., Fangmeier, T. (2024). Processing of prosodic cues of uncertainty in autistic and non-autistic adults: a study based on articulatory speech synthesis. *Frontiers in Psychiatry, Sec. Autism*, Vol. 15, 14 October 2024, https://doi.org/10.3389/fpsyt.2024.1347913

Bellinghausen. S., Riedel, A. (2025). Characteristics of Autistic and Neurotypical Speech and Automated Detection: A Systematic Review of Linguistic Analysis. Poster presented at *Wissenschaftliche Tagung Autismus Spektrum 2025*, Heidelberg, Germany, 17-18th of April 2025.

Black, M.P., Bone, D., Williams, M.E., Gorrindo, P., Levitt, P., Narayanan, S. 2011. The USC CARE corpus: child-psychologist interactions of children with autism spectrum disorders. In *Proceedings of Interspeech 2011*, Florence, Italy, 1497-1500, doi: 10.21437/Interspeech.2011-314

Chi, N.A., Washington, P., Kline, A., Husic, A., Hou, C., He, C., Dunlap, K., Wall, D.P. 2022. Classifying Autism From Crowdsourced Semistructured Speech Recordings: Machine Learning Model Comparison Study. JMIR Pediatr Parent. 2022 Apr 14;5(2):e35406. doi: 10.2196/35406. PMID: 35436234; PMCID: PMC9052034.

Cho, S., Liberman, M., Ryant, N., Cola, M., Schultz, R.T., Parish-Morris, J. 2019. Automatic detection of autism spectrum disorder in children using acoustic and text features from brief natural conversations. In *Proceedings of Interspeech 2019*, Graz, Austria, 2513–2517 doi: 10.21437/Interspeech.2019-1452

Diehl, J.J., Paul, R. 2012. Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Autism Spectr Disord*. 6:123–34. doi: 10.1016/j.rasd.2011.03.012

Eberhardt-Juchen, M. 2023. Sprach- und Kommunikationsentwicklung bei Autismus einschließlich Besonderheiten der Sprache und Kommunikation bei Autismus. In S. Lindmeier, S. Sallat, K. Ehrenberg, *Sprache und Kommunikation bei Autismus*. Stuttgart, Verlag W. Kohlhammer, 19-34.

Eni, M., Dinstein, I., Zigel, Y. 2023. Comparing Hand-Crafted Features to Spectrograms for Autism Severity Estimation. In *Proceedings of Interspeech 2023*, Dublin, Ireland, 4154-4158, doi: 10.21437/Interspeech.2023-658

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*, 2017 Mar, 10(3), 384-407. doi: 10.1002/aur.1678

Ganek, H., Eriks-Brophy, A. 2018. Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. *J Commun Disord.* 2018 Mar-Apr; 72:77-85. doi: 10.1016/j.jcomdis.2017.12.005. Epub 2017 Dec 16. PMID: 29402382.

Geurts, H.M., Embrechts, M. 2008. Language profiles in asd, sli, and adhd. *Journal of autism and developmental disorders*, 38:1931–1943. doi: 10.1007/s10803-008-0587-1

Geurts, B., Kissine, M., van Tiel B. 2019. Pragmatic reasoning in autism. In: K. Morsanyi, R.M.J. Byrne RMJ: *Thinking, reasoning and decision making in autism*. Routledge, London, 113–34.

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 Res*. (2015) 8:153–63. doi: 10.1002/aur.1432

ICD-11 ‘autism’. 2025). https://icd.who.int/browse/2025-01/mms/en#437815624 [last access: 2nd of April 2025]

ISCA. 2025. International Speech Association Archive. https://www.isca-archive.org/ [last access: 2nd of April 2025]

Kamp-Becker, I., Bölte S. 2014. Autismus. München, Basel, Ernst Reinhard Verlag.

Kiss, G., van Santen, J.P.H., Prud'hommeaux, E., Black, L.M. 2012. Quantitative analysis of pitch in speech of children with neurodevelopmental disorders. In *Proceedings of Interspeech 2012*, Portland, OR, USA, 2012 Sep 9-13. 1343–46.

Lahiri, R., Feng, T., Hebbar, R., Lord, C., Kim, S.H., Narayanan, S. 2023. Robust Self Supervised Speech Embeddings for Child-Adult Classification in Interactions involving Children with Autism. In *Proceedings of Interspeech 2023*, Dublin, Ireland, 3557-3561, doi: 10.21437/Interspeech.2023-1447

Lee, H.-y., Hu, T.-y., Jing, H., Chang, Y.-F., Tsao, Y., Kao, Y.-C., Pao, T.-L. 2013. Ensemble of machine learning and acoustic segment model techniques for speech emotion and autism spectrum disorders recognition. In *Proceedings of Interspeech 2013*, Lyon, France, 215-219, doi: 10.21437/Interspeech.2013-70

Li, J., Chen, Z., Li, G., Ouyang, G., Li, X. 2022. Automatic classification of asd children using appearance-based features from videos. *Neurocomputing*, 470:40–50, doi: 10.1016/j.neucom.2021.10.074

Lindmeier, S., Sallat, S., Ehrenberg. K. 2023. *Sprache und Kommunikation bei Autismus*. Stuttgart, Verlag W. Kohlhammer.

Ma, W., Xu, L., Zhang, H., Zhang, S. 2024. Can Natural Speech Prosody Distinguish Autism Spectrum Disorders? A Meta-Analysis. *Behav Sci*, 2024 Jan 26, 14(2):90, doi: 10.3390/bs14020090. PMID: 38392443; PMCID: PMC10886261

MacFarlane, H., Salem, A.C., Chen, L., Asgari, M., Fombonne, E. 2022. Combining voice and language features improves automated autism detection. *Autism Research 2022*, 15(7):1288-1300, doi: 10.1002/aur.2733. Epub 2022 Apr 23. PMID: 35460329; PMCID: PMC9253091.

Markowetz, R. 2020. *Schüler mit Autismus-Spektrum-Störung im inklusiven Unterricht. Praxistipps für Lehrkräfte*. München, Ernst Reinhardt Verlag.

Neitzel, I., Tuschen, L., Ritterfeld, U. 2023. Automatisierte Sprach- und Sprechanalyse in Forschung und Diagnostik bei Autismus-Spektrum-Störungen: Potenziale und Barrieren. *Sprache · Stimme · Gehör 2023*, 47(02): 84-88 doi: 10.1055/a-2018-2972

Philofsky, A., Fidler, D.J., Hepburn, S. 2007. Pragmatic language profiles of school-age children with autism spectrum disorders and williams syndrome. *American Journal of Speech-Language Pathology*, 16(4):368-80. doi: 10.1044/1058-0360(2007/040). PMID: 17971496; PMCID: PMC4426201.

Probol, N., Mieskes, M. 2024. Autism Detection in Speech – A Survey. In *Findings of the Association for Computational Linguistics: EACL 2024*, Malta, March 17-22, 1115–1125. doi: 10.48550/arXiv.2402.12880

Prud'hommeaux, E., Rouhizadeh, M. 2012. Automatic detection of pragmatic deficits in children with autism. In *Workshop Child Comput Interact*. 2012 Sep 14; 2012, 1-6. PMID: 28691126; PMCID: PMC5500165.

PubMed. 2025. Pub Med Website. https://pubmed.ncbi.nlm.nih.gov/ [last access: 2nd of April 2025]

Riedel A. (2015). Sprachpragmatik bei Autismus-Spektrum-Störungen. In 8. Newsletter, Wissenschaftliche Gesellschaft Autismus-Spektrum-Störung (2015). https://silo.tips/download/newsletter-sprachpragmatik-bei-autismus-spektrum-strungen-mai-nr-8-dr-med-dr-phi

Riedel, A., Biscaldi-Schäfer. 2025, in press. *Sprache bei Menschen im Autismus-Spektrum. Interdisziplinäre Perspektiven*. Stuttgart, Verlag W. Kohlhammer.

Shriberg, L.D., Paul, R., McSweeney, J.L., Klin, A.M., Cohen, D.J., Volkmar, F.R. 2001. Speech and prosody characteristics of adolescents and adults with high functioning autism and Asperger syndrome. J Speech Lang Hear Res., 44:1097–115. doi: 10.1044/1092-4388(2001/087). PMID: 11708530.

Theunissen, G. 2020. Autismus – das neue Verständnis aus der Außenansicht in Anlehnung an Vorstellungen von Betroffenen. In G. Theunissen: *Autismus verstehen. Außen- und Innenansichten*, Stuttgart, Verlag W. Kohlhammer, 21-101.

Vogindroukas, I., Stankova, M., Chelas, E.-N., Proedrou, A. 2022. Language and speech characteristics in autism. *Neuropsychiatric Disease and Treatment*, 2367–2377. doi: 10.2147/NDT.S331987. PMID: 36268264; PMCID: PMC9578461.

Volden, J., Coolican, J., Garon, N., White, J., Bryson, S. 2009. Brief report: Pragmatic language in autism spectrum disorder: Relationships to measures of ability and disability. *Journal of autism and developmental disorders*, 39:388–393. doi: 10.1007/s10803-008-0618-y. Epub 2008 Jul 15. PMID: 18626760.

Wetherby, A.M., Watt, N., Morgan, L., Shumway, S. Social communication profiles of children with autism spectrum disorders late in the second year of life. *J Autism Dev Disord.*, 2007 May, 37(5):960-75. doi: 10.1007/s10803-006-0237-4

WHO - World Health Organization Website. 2025 .https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders [last access: 2nd of April 2025]

Xu, D., Gilkerson, J., Richards, J., Yapanel, U., Gray, S. 2009. Child vocalization composition as discriminant information for automatic autism detection. *Annu Int Conf IEEE Eng Med Biol Soc. 2009*; 2518-22. doi: 10.1109/IEMBS.2009.5334846

Xu, A., Hebbar, R., Lahiri, R., Feng, T., Butler, L., Shen, L., Tager-Flusberg, H., Narayanan, S. 2023. Understanding Spoken Language Development of Children with ASD Using Pre-trained Speech Embeddings. *Proceedings of Interspeech 2023*, 4633-4637, doi: 10.21437/Interspeech.2023-1273

Yang, Y., Cho, S., Covello, M., Knox, A., Bastani, O., Weimer, J., Dobriban, E., Schultz, R., Lee, I., Parish-Morris, J. 2023. Automatically Predicting Perceived Conversation Quality in a Pediatric Sample Enriched for Autism. In *Proceedings of Interspeech 2023*, Dublin, Ireland, 4603-4607, doi: 10.21437/Interspeech.2023-2251

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M.S., Saxena, S., Yusuf, A., Shih, A., Elsabbagh, M. 2022. Global prevalence of autism: A systematic review update. *Autism Res.* 2022 May; 15(5):778-790. doi: 10.1002/aur.2696. Epub 2022 Mar 3. PMID: 35238171; PMCID: PMC9310578.