

RESEARCH ARTICLE

Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study

Joseph C. Y. Lau¹, Shivani Patel¹, Xin Kang^{2,3,4,5}, Kritika Nayar¹, Gary E. Martin⁶, Jason Choy², Patrick C. M. Wong^{2,3}, Molly Losh^{1*}

1 Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois, United States of America, **2** Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong S.A.R., China, **3** Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong S.A.R., China, **4** Research Centre for Language, Cognition and Language Application, Chongqing University, Chongqing, China, **5** School of Foreign Languages and Cultures, Chongqing University, Chongqing, China, **6** Department of Communication Sciences and Disorders, St. John's University, Staten Island, New York, United States of America

* m-losh@northwestern.edu



OPEN ACCESS

Citation: Lau JCY, Patel S, Kang X, Nayar K, Martin GE, Choy J, et al. (2022) Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study. *PLoS ONE* 17(6): e0269637. <https://doi.org/10.1371/journal.pone.0269637>

Editor: Carla Pegoraro, PLOS, UNITED KINGDOM

Received: February 14, 2022

Accepted: May 24, 2022

Published: June 8, 2022

Copyright: © 2022 Lau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The minimal dataset and analytic scripts are available from the OSF database (<https://osf.io/9ta65/>).

Funding: This research was supported by grants from the National Institutes of Health (USA: R01DC010191, R03DC018644, PI: ML), Health and Medical Research Fund (Hong Kong: 02130846, PI: PW), Global Parent Child Resource Centre Limited (to PW), and Dr Stanley Ho Medical Development Foundation (to PW). The funders had no role in study design, data collection and

Abstract

Differences in speech prosody are a widely observed feature of Autism Spectrum Disorder (ASD). However, it is unclear how prosodic differences in ASD manifest across different languages that demonstrate cross-linguistic variability in prosody. Using a supervised machine-learning analytic approach, we examined acoustic features relevant to rhythmic and intonational aspects of prosody derived from narrative samples elicited in English and Cantonese, two typologically and prosodically distinct languages. Our models revealed successful classification of ASD diagnosis using rhythm-relative features within and across both languages. Classification with intonation-relevant features was significant for English but not Cantonese. Results highlight differences in rhythm as a key prosodic feature impacted in ASD, and also demonstrate important variability in other prosodic properties that appear to be modulated by language-specific differences, such as intonation.

Introduction

Differences in prosody are observed in many individuals with autism spectrum disorder (ASD) [1, 2]. Although they may manifest variably across individuals and across different language contexts (e.g., [3–5], also see [2, 6] for reviews), prosodic differences have been considered a central feature of communication profile of ASD [7–10]. Speech prosody involves the use of rhythm (i.e., variation of regularity in loudness and speed) and intonation (i.e., variation of voice pitch) [6] to encode grammatical information, represent pragmatic information, as well as to express speaker intent and emotion [11–13]. Differences in prosody can significantly undermine social and communicative competence by disrupting communication of this linguistic and meta-linguistic information and contrasts (e.g., parts of speech differences [e.g., CONtent vs. conTENT]; a sarcastic vs. sincere statement; joy vs. dislike), and is therefore

analysis, decision to publish, or preparation of the manuscript.

Competing interests: I also declare that I have read the journal's policy and the authors of this manuscript have the following competing interests: Patrick C.M. Wong, declares that he is an owner of a startup company supported by a Hong Kong Government technology startup scheme. The research reported here is not associated with the company. This does not alter our adherence to PLOS ONE policies on sharing data and materials. The other authors declare no competing interests exist. All authors have approved the manuscript and have contributed significantly for this research. The minimal dataset and analytic scripts are available from the OSF database (<https://osf.io/9ta65/>).

highly clinically significant. Additionally, subtle differences in prosody have also been reported among clinically unaffected first-degree relatives, and could constitute a marker of underlying genetic liability to ASD [14].

Although a number of studies have provided important insights into the characteristics of speech prosody in individuals with ASD, little is known concerning the underlying causes of prosodic differences in ASD. Moreover, most studies have concentrated on studying prosody in English-speaking groups [2, 6]. Language is fundamentally cultural, and how specific aspects of prosody are used can vary substantially across languages [15]. Studying language components, such as prosody, across typologically and prosodically distinct languages is therefore critical to address the variability of ASD symptomatology, and its underlying mechanisms.

Since the earliest delineations of ASD [16], prosodic differences have been described in subjective reports as a characteristic of ASD (see [1] for a review). Subsequent studies applied more objective methodologies examining the physical (acoustic) profile of prosody in ASD, including the fundamental frequency (f_0), duration, and periodicity of speech sounds—acoustic properties which are correlates of fundamental aspects of prosody, namely intonation and rhythm. Findings have largely converged, showing differences in intonational properties of prosody (e.g., pitch), in ASD [2, 6]. Evidence suggests that individuals with ASD often demonstrate overall higher f_0 (the acoustic correlate of pitch) [17], and larger f_0 range at the utterance level [17–21], at the syllabic level [22], at the utterance final position [14], and specifically when using focus to highlight new information [23]. In rhythmic aspects of prosody, speech produced by individuals with ASD also demonstrated less of a distinction in duration between stress and unstressed syllables [22], slower speech rate [14], as well as greater intensity and longer phrase durations [20]. Although these prosodic features do not usually result in unintelligible speech, they are nevertheless perceptible [14] and can contribute to an impression of “oddness” reported by listeners [20, 24, 25].

Traditionally, speech prosody is categorized into affective and linguistic categories [26]. Affective prosody is used to express emotion and communicative intent, both central to social interactions, through non-linguistic elements of speech [27]. Despite some cross-cultural variabilities, fundamental emotions and basic communicative intents expressed through prosody are typically similarly understood by listeners across different linguistic and cultural backgrounds [28, 29], and may even have evolutionary homologies in other mammals [30–32]. In contrast, linguistic prosody is used to represent lexical and grammatical elements of language, and may vary drastically across languages [15, 33]. In particular, while pitch signifies sentence intonation (e.g., to convey a statement vs. a question) in English, pitch also conveys *lexical* meaning in tone languages such as Cantonese. For example, in Cantonese, a syllable /ji/ means ‘to cure’ when produced with a high-level pitch pattern but means ‘two’ when produced with a low-level pitch pattern.

Prosodic expression in individuals with ASD speaking a tone language has been scarcely investigated. One study examined the f_0 and duration of the five Thai lexical tones produced by native speakers with ASD [34]. A separate study of Cantonese-speaking individuals with ASD examined f_0 measures of sentence final particles (SFPs), which are crucial linguistic markers for conveying pragmatic information in the language [35]. Another study examined lexical tone imitation of Cantonese- vs. Mandarin-speaking children with ASD [36]. These studies revealed findings similar to those reported in English-speaking individuals with ASD, such that the individuals with ASD exhibited higher f_0 [34] and larger standard deviation (i.e., larger variability) of f_0 [35, 36]. Some language-specific results were also observed, including shorter lexical tone duration in the ASD group [34]. A positive correlation between the number of SFP types produced by individuals with ASD and the standard deviation of F_0 was also found [35], suggesting the more diverse the SFP type an individual produced, the more pitch

variation could be realized at the utterance-final position which does not alter core sentential content. Some language-specific findings were not consistent with those reported in studies of English. Specifically, while larger f_0 range has been repeatedly reported in English-speaking ASD groups [17–19, 21, 22], Thongseitratch and colleagues [34] found a lower f_0 range in the Thai-speaking ASD group. A lower f_0 range associated with ASD was also found in a study of Japanese speakers [3]. These mixed results from languages other than English suggest that while certain acoustic characteristics of speech may be impacted across languages, typological differences across languages may also result in important language-specific prosodic characteristics associated with ASD. Such differences are not only essential in informing the understanding of prosodic differences as a core feature of ASD, but may also help to optimize culturally sensitive diagnostic and intervention practices.

Taking a cross-linguistic approach, the present study compared prosodic characteristics of speech from individuals with and without ASD across English and Cantonese, two typologically and prosodically distinct languages. English is an Indo-European Germanic non-tone language that belongs to the rhythmic group of stress-timed languages [33], whereas Cantonese is a Sino-Tibetan Sinitic tone language which is syllable-timed [37]. In the general population, even in languages with such distinct properties, considerable cross-linguistic commonalities have been demonstrated in prosody [28, 29, 38–40], particularly in the affective aspect of prosody that is most reflective of differences in ASD (e.g., emotion and intention expression and recognition). Important cross-linguistic variability has also been observed, especially in linguistic prosody [15, 33]. Indeed, prior cross-linguistic comparisons between prosodic characteristics of ASD have been reported in languages which are typologically and prosodically related (e.g., English and Danish, both Germanic languages [41], and Cantonese and Mandarin, both Sinitic languages [36]). However, studying prosody in ASD across languages with more distinct properties, such as English and Cantonese, may help to further reveal core prosodic differences that are robustly expressed across languages, as well as those that are differentially impacted, suggesting phenotypic malleability that could be important to consider in speech and language interventions.

We implemented a series of novel machine learning (ML)-based analytics to examine prosodic characteristics of ASD both within and across the two languages. ML capitalizes on the multivariate nature of acoustics features (e.g., series of features that vary in time or frequency domains rather than single summed or averaged values) that richly represents the dynamicity of speech prosody, and have been proven to be able to delineate the acoustic characteristics of the prosodic profile of ASD [6, 42–44]. We focused on two classes of acoustic features, namely those relevant to 1) *rhythm*, and 2) *intonation*, considered to be the two core elements of speech prosody [12]. A series of supervised ML models were trained to make classifications of diagnosis (ASD vs. typical development [TD]) using a series of multivariate acoustic features associated with intonation and rhythm, derived from utterance samples of a structured narrative task elicited separately in English and Cantonese. [45–47]. ML classification models have been used successfully to classify individuals with ASD from those with TD using a variety of social-behavioral and neurocognitive measures (see [48] for a review). ML classification models have also proven to be effective in classifying diagnostic status of various psychiatric, cognitive, and speech-language impairments using speech acoustic features [49–51]. In particular, a recent ML study successfully classified ASD/TD using f_0 -based acoustic features derived from recordings of individual words elicited in a picture-naming task [44]. Capitalizing on the power of the ML classification approach, our ML models examined which class(es) of acoustic features (*rhythm* and/or *intonation*) derived from larger temporal windows (i.e., from utterances elicited in a more dynamic language narrative task) could reliably classify individuals with ASD vs. individuals with TD, both within and across English and Cantonese.

Materials and methods

Participants

This study included participants from two languages groups, consisting of native speakers of American English (henceforth, English group) and Hong Kong Cantonese (henceforth, Cantonese group). The English group included 55 individuals with ASD (English ASD group) and 39 individuals with TD (English TD group). The Cantonese group included 28 individuals with ASD (Cantonese ASD group) and 24 individuals with TD (Cantonese TD group).

We acknowledge the variability in preferences related to person-first and identity-first terminology (i.e., individuals with ASD versus autistic individuals). We do not endorse one style over the other. For the purpose of this manuscript, to ensure that language is parallel when referencing both participant groups, we remain consistent with the following terminology: individuals with autism spectrum disorder (ASD) and individuals with typical development (TD).

Participants from the English group were recruited in the United States through a larger family-genetic study of ASD, which included individuals with ASD, their parents, and respective controls. Participants from the Cantonese group were recruited in Hong Kong using advertisements posted on social media platforms (e.g., Facebook) and directly sent to schools and organizations with existing populations of individuals with ASD, as well as from employment programs particularly designed for adults previously diagnosed with Asperger syndrome or high-functioning Autism. Informed assent/consent was obtained from all participants and guardians (as applicable), and procedures were approved by the respective institution's ethics committee. All procedures were in accordance with ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Only participants having no reported history of brain injury, major psychiatric disorder, known genetic syndrome, or neurodevelopmental disorder (other than ASD) were included. Participants with TD in both language groups were screened for personal or family history of ASD or related genetic disorders (e.g., fragile X syndrome). ASD status was confirmed with research reliable administration and scoring of the Autism Diagnostic Observation Schedule 2nd Edition (ADOS-2). ADOS-2 Overall, Social Affect, and Restricted and Repetitive Behaviors (RRB) calibrated severity scores were used to determine ASD severity [52].

To assess IQ, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999), Wechsler Adult Intelligence Scale-Third or Fourth Editions (WAIS; Wechsler 1997, 2008), or the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler 2003) were administered to participants in the English group. The Test of Nonverbal Intelligence, Fourth Edition (TONI-4) was administered to participants in the Cantonese group.

Narrative elicitation

Participants were asked to narrate (in their respective languages) the 24-page wordless picture book, *Frog, Where Are You?* [53]. The book presents a story about a boy and his dog, who are searching for the boy's missing pet frog. This book has been used extensively in studies of narrative discourse in ASD and other neurodevelopmental disabilities [14, 54–58] and in cross-linguistic work [59]. Participants were asked to narrate the story as each page of the book was presented to participants on a computer monitor, while their narrations were audio recorded. No examples were provided to the participant. The recordings were first transcribed and segmented into individual utterances, defined by natural pauses. Twenty utterances from participants (English ASD $n = 33$; English TD $n = 33$; Cantonese ASD $n = 24$; Cantonese TD $n = 24$,

Table 1. Demographic information.

	ASD (Cantonese)	TD (Cantonese)	ASD (English)	TD (English)
	M (S.D.) Range	M (S.D.) Range	M (S.D.) Range	M (S.D.) Range
Males: females (Count)	19:9	17:7	29:4	15:18
Chronological age	17.83 (9.24) 8–32	18.88 (8.70) 8–31	15.96 (7.38) 6–35	19.31 (5.35) 12–32
IQ	108.54 (10.92) 84–127	115.08 (10.13) 92–128	<i>104.13 (14.34)</i> 73–131	<i>111.03 (13.79)</i> 79–143
ADOS-2 Total Severity Score	6.00 (2.33) 3–10		7.02 (1.85) 3–9.5	

Bold: Significant differences as per t-tests ($p < 0.05$) between the ASD and TD groups within the respective language group. *Italics:* Marginal differences as per t-tests ($0.05 > p < 0.1$) between the ASD and TD groups within the respective language group. ASD: Autism Spectrum Disorder, M: Mean, S.D.: standard deviation, TD: typical development

<https://doi.org/10.1371/journal.pone.0269637.t001>

Table 1) were selected for subsequent analyses (see §1.1 in S1 File for further description of acoustic data collection and processing).

The narrative data elicited from the English group were the focus of a prior published study [14], examining syllable-level acoustic measures that were not included in the present analyses. The English participant narrative speech samples were included in this study for cross-linguistic comparison only focused on utterance-level acoustic measures that have not been previously studied. In contrast to recent studies in both English [14] and Cantonese [35] reporting prosodic differences associated with ASD at language-specific syllabic-level units (final-syllable excursion for English; SFP for Cantonese), the current study focused on the *utterance* level. In contrast to syllabic-level units, the *utterance* is associated with the intonational phrase, the most fundamental unit of prosody argued to be universal across all languages [60–63]. The ML approach allowed us to perform classifications using multivariate acoustic features representing the prosodic dynamics over the entire duration of utterances (described in the following section), therefore enabling us to examine, at the most fundamental prosodic unit common among English and Cantonese, the extent to which patterns of speech prosodic differences in ASD similarly manifest across different languages.

Acoustic feature extraction

Two classes of multivariate acoustic features were extracted from utterances in the narrative samples utilizing the MATLAB Audio Toolbox and MATLAB scripts provided by prior studies [45–47, 64].

1. **Speech Rhythm:** Speech rhythm is traditionally considered as durational variations across syllables in an utterance that signal linguistic and affective properties [12]. The temporal envelope of speech signals, in general, represents the evolution of the speech signal waveform amplitude over time that displays temporal regularities correlating to the syllabic rhythm of the signal [64]. Such rhythmicity, especially those of 2–8 Hz, is crucial in the neural processing of intelligible speech as it aligns with brain areas in the spoken language processing pathway that oscillate at a similar rate [65]. Three measures, namely 1) envelop spectrum (ENV), 2) intrinsic mode functions (IMF), and 3) temporal modulation spectrum (TMS) [45–47] were derived from all utterances of each participant to comprehensively capture aspects of speech rhythm represented in the temporal envelope. An overall of 8640 these rhythm-relevant features were extracted across the 20 utterances from each participant.

2. **Intonation:** Speech intonation, by definition [61], refers to the variation of voice pitch across time. The major acoustic correlate of pitch is the fundamental frequency (f0), which is a simple yet robust index of speech intonation that has repeatedly been implicated in the prosodic characteristics of ASD [17–21, 44]. In contrast to univariate measures (single values) of f0 used in prior studies [17–21], often computed by taking the mean or variance over multiple temporal windows or across speech samples, we focused on entire f0 contours, i.e., a series of time varying f0 values (i.e., multivariate) across the entire duration of each individual utterance. The f0 contour represents essential elements of both linguistic prosody and affective prosody at the sentence/utterance level [11, 61, 66]. In particular, the f0 contour has been the target of study in both empirical and theoretical work of intonational phonology [61, 63], as it encompasses the dynamicity of speech intonation which varies across time within the most fundamental utterance-sized linguistic unit known as the intonational phrase. Therefore, intonation-relevant features examined in this study were comprised of time-varying f0 values derived from narrative utterances. From each utterance, 20 f0 values from a time-normalized f0 contour were extracted and further concatenated across all 20 utterances of a participant, resulting in 400 features for each participant.

Further technical descriptions of the acoustic feature extraction procedures are presented in §1.2 in [S1 File](#).

Machine learning classification

In two sets of ML models, we trained a series of linear support vector machine (SVM) classifiers to classify participants with ASD from those with TD using principal components derived from the multivariate acoustic features extracted from participants' narrative samples. The SVM performed classifications by finding a hyperplane from a multidimensional space that divided data points from the principal components according to participants' diagnosis.

The SVM was chosen because it is well suited to handle data of high-dimension, and the types of speech acoustic features implicated in ASD (e.g., f0-based features) in particular [44]. The linear kernel was chosen due to its effectiveness in handling datasets with a small sample size [67]. Widely chosen in studies with datasets where the number of features often even exceeds that of samples (e.g., neuroimaging [68] and gene expression studies [69]), the linear SVM is preferable to non-linear kernels because theoretically it is always possible to find a linear decision boundary that separates data, in spite of high data dimensionality and small sample size [68].

Several procedures were performed in the classification pipeline to further avoid overfitting and optimistic bias due to limited sample size and the high-dimensionality of data [70] (see §1.3 in [S1 File](#) for technical details of procedures in the machine learning classification pipeline). Classifications were performed using a repeated 10-fold cross-validation procedure, in which data reduction (into principal components) and hyperparameter tuning were performed in a nested fashion. Classification performance was quantified as the Area Under the Curve (AUC) of a receiver operating characteristics (ROC) curve computed based on the probability vector of the predicted labels across all cross-validation folds. The accuracy, sensitivity, and specificity values of the classification were also recorded. A permutation approach was used to estimate the statistical significance of each series of classification with the AUC values.

In Model 1, classifications were performed separately on English and Cantonese samples. The comparison between classification performance patterns across the English and Cantonese classifications allowed us to both identify the specific acoustic characteristics associated

with prosody in ASD within these two languages, as well as determine whether such characteristics were consistent or different between the two languages. In Model 2, ASD vs. TD classifications were performed on an English and Cantonese-combined dataset in order to identify aspects of prosody that were present in ASD narrative samples across the two languages. In each model, two sets of SVM classifications were performed using the principal components derived from *rhythm*- and *intonation*-relevant features respectively.

Results

Model 1 performed classifications between ASD and TD diagnosis using *rhythm*- (ENV, IMF, and TMS) or *intonation*-relevant (f0 contours) features, on English and Cantonese samples respectively. Model statistics of all classifications are presented in Table 2. The AUC values of all SVM classifications in Model 1 are presented in Fig 1 (left and middle). The classifications with *rhythm*-relevant features were significant for both English (median AUC = 0.900, $p < 0.001$ —the median AUC corresponded to an accuracy [ACC] of 0.819, sensitivity [SENS] of 0.788, and specificity [SPEC] of 0.849), and Cantonese (median AUC = 0.962, $p < 0.001$; ACC = 0.880, SENS = 0.917, SPEC = 0.833). In contrast, classifications with *intonation*-relevant features were significant only for English (median AUC = 0.695, $p = 0.007$; ACC = 0.683, SENS = 0.758, SPEC = 0.606) but not Cantonese (median AUC = 0.620, $p = 0.507$; ACC = 0.605, SENS = 0.667, SPEC = 0.542). A post-hoc analysis (presented in §2 in S1 File) was further performed to rule out gender and age as potential confounding factors in the f0-based classification in English, given differences in gender and age observed between the English (but not Cantonese) groups.

Model 2 further examined the classifications between ASD and TD diagnosis using *rhythm*- or *intonation*-relevant features, using a dataset combining both English and Cantonese samples. The AUC values of all SVM classifications in Model 2 are presented in Fig 1 (right). The classifications using *rhythm*-relevant features was significant (median AUC = 0.886, $p < 0.001$; ACC = 0.835, SENS = 0.790, SPEC = 0.877), whereas the classification using *intonation*-relevant features was near chance level (median AUC = 0.559, $p = 0.509$; ACC = 0.566, SENS = 0.632, SPEC = 0.509).

Confusion matrices of all classifications are presented in Fig 2. In general, all classifications using rhythmic features showed comparably high SENS (accuracy in predicting ASD cases) and SPEC (accuracy in predicting TD cases) rate. SPEC is higher in Model 1 on English and Model 2, whereas SENS was higher for Model 1 on Cantonese. In contrast, although SPEC was low in the statistically significant Model 1 on English using intonational features, good SENS was observed. The non-significant classifications using intonational features had generally low SENS and SPEC.

Table 2. Model statistics.

Model	Language	Features	Median AUC	ACC	SENS	SPEC
1	English	Rhythm	0.900***	0.819	0.788	0.849
		Intonation	0.695**	0.683	0.758	0.606
	Cantonese	Rhythm	0.962***	0.880	0.917	0.833
		Intonation	0.620	0.605	0.667	0.542
2	English & Cantonese	Rhythm	0.886***	0.835	0.790	0.877
		Intonation	0.559	0.566	0.632	0.509

Model median area-under-the-curve (AUC) and associated accuracy (ACC), sensitivity (SENS), and specificity (SPEC)

***Permutation $p < .001$.

<https://doi.org/10.1371/journal.pone.0269637.t002>

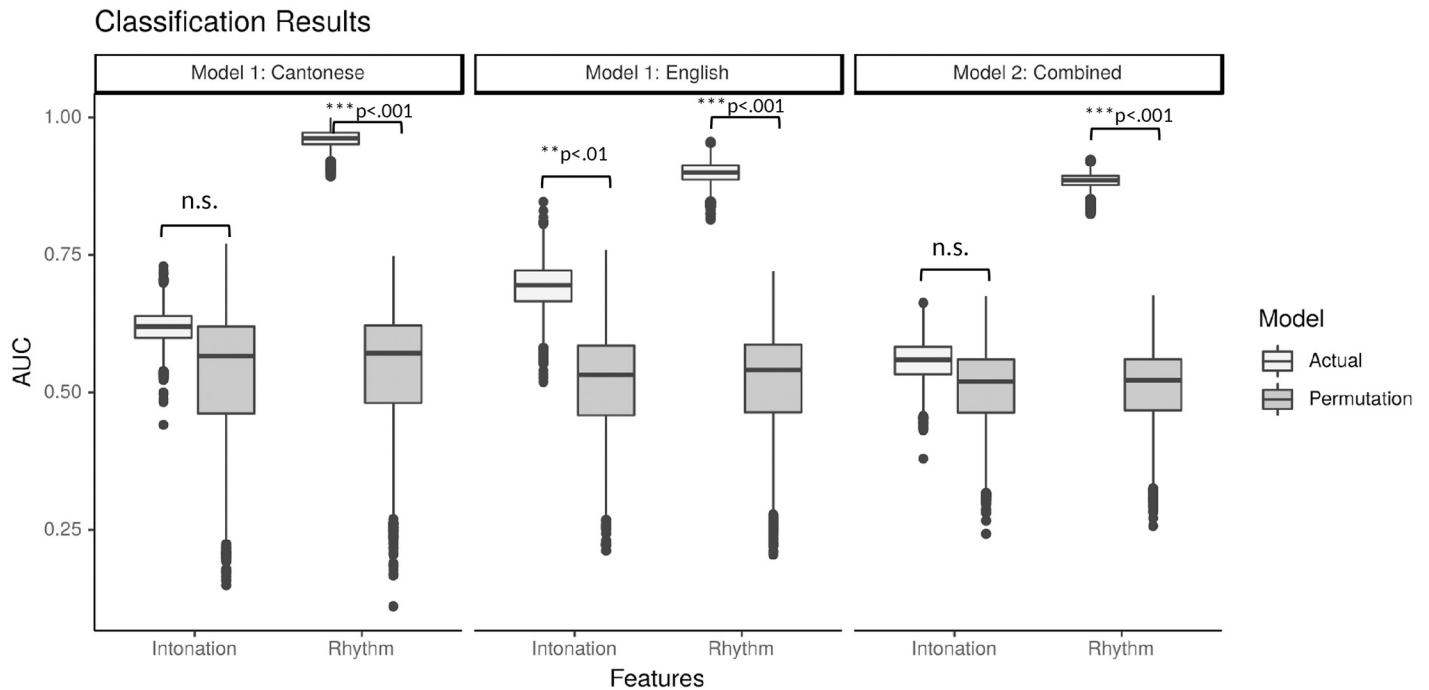


Fig 1. Machine learning classification results. Machine learning classification results displayed in boxplots of Area-Under-the-Curve values across 5001 iterations and permutations.

<https://doi.org/10.1371/journal.pone.0269637.g001>

Discussion

The present study applied machine learning (ML)-based analyses to acoustic features of speech in ASD, in an attempt to identify prosodic differences in ASD that span two typologically distinct languages, as well as those that may be shaped by specific linguistic and cultural influences. Capitalizing on the power of ML algorithms for performing classifications using high dimensional data [6, 42–44], we examined acoustic features representing intonational and rhythmic aspects of prosody across time or frequency domains. Findings from ML-based algorithms demonstrated that acoustic features can be used to reliably classify ASD vs. TD group status in both English and Cantonese. Results demonstrate the value of moving beyond more reductionist approaches (e.g., averaging or picking extreme values) often used in traditional univariate acoustic analyses on pitch, stress, and speech rate [2], to capture the dynamicity of speech prosodic profiles, and point to differences in prosodic features that may be robust characteristics of the ASD speech and language phenotype in multiple languages.

Specifically, results of Model 1 indicated that prosodic features in the rhythm of speech (i.e., the envelop spectrum [ENV], intrinsic mode functions [IMF], and temporal modulation spectrum [TMS]—representations of temporal regularities of speech signal waveforms correlating to the syllabic rhythm) contained crucial information captured by the ML algorithm to differentiate individuals with ASD from controls. Such findings complement the many rich, descriptive studies of prosody conducted previously, that described differences in stress pattern, speech rate, and loudness that could be attributed to ASD [4, 71–75]. It is notable that the ML algorithms identified in this study appeared to capture patterns from acoustic features that were representative of these prosodic characteristics, paralleling such descriptive observations by clinicians and researchers. Results of Model 1 also converge with using objective acoustic measures to characterize speech rhythm in ASD, which have strongly implicated that acoustic

		Model 1: English		Model 1: Cantonese		Model 2: Combined			
Rhythm	ASD	131536 (79.7%)	18942 (11.5%)	113467 (94.5%)	18146 (15.1%)	221900 (77.8%)	43988 (15.4%)	Prediction	
	TD	33497 (20.3%)	146091 (88.5%)	6557 (5.5%)	101878 (84.9%)	63157 (22.2%)	241069 (84.6%)		
Intonation	ASD	118311 (71.7%)	72840 (44.1%)	72728 (60.6%)	49161 (41.0%)	178031 (62.5%)	147758 (51.8%)		
	TD	46722 (28.3%)	92193 (55.9%)	47296 (39.4%)	70863 (59.0%)	107026 (37.5%)	137299 (48.2%)		
		ASD	TD	ASD	TD	ASD	TD		
		True Class							

Fig 2. Confusion matrices. Confusion Matrices of machine learning classifications in Model 1 (English and Cantonese) and Model 2 (Combined), aggregated across all 5001 iterations of cross-validation in each classification. Blue hues: correct predictions; Red hues: incorrect predictions.

<https://doi.org/10.1371/journal.pone.0269637.g002>

measures of speech rate and lexical stress based on stressed syllable duration, f_0 , as well as cross-syllabic durational variability (e.g., the Normalized Pairwise Variability Index) varied as a function of ASD diagnosis [14, 22].

Perhaps a more revealing aspect of Model 1 is its cross-linguistic design, examining both English and Cantonese utterance samples. Relatively few studies [3, 34–36] have examined prosody in ASD in languages prosodically and typologically distinct from English [41]. All of these studies focused only on pitch, and only one [36] applied a cross-linguistic design to allow direct comparisons of prosody across multiple languages, in the same speech sampling context. The ML algorithm employed in Model 1 was able to classify ASD diagnosis using rhythm-relevant acoustic features (ENV, IMF, and TMS) derived from both English and Cantonese, revealing strong performance for classification (median AUCs ~ 0.9) in both languages respectively. Given the considerable typological differences between English and Cantonese, it is notable that the same type of acoustic features derived from English and Cantonese produced reliable ASD/TD classifications, suggesting that speech rhythm is an important feature of the prosodic profile of ASD that is evident in multiple languages. In Model 2, with datasets combined, the classification only using the rhythm-relevant acoustic features was also robust (median AUC > 0.8), particularly considering the large linguistic and acoustic variability introduced to the dataset by collapsing the two languages (see §3 in S1 File for a supplementary ML analysis demonstration of systematic cross-linguistic differences present in our rhythm-relevant features). These results suggest that there are rhythmic characteristics of prosody associated with ASD in both English and Cantonese, that potentially represent aspects of a meaningful prosodic profile in ASD that is robustly expressed across the two typologically distinct languages. These results also complement prior cross-linguistic work reporting common rhythmic-related characteristics (i.e., pause length) across languages, i.e., English and Danish [41]. Identification of such a constellation of acoustic features that can reliably predict ASD diagnostic status, across different languages, may hold significant potential for contributing to diagnostic and intervention practices, as well as studies to understand the basis of potential language-related impairments in ASD.

Such cross-linguistic commonalities in rhythmic characteristics of prosody associated with ASD may partly stem from the crucial role of prosody in social communication that is partly invariant across cultures. Speech prosody, in general, demonstrates strong similarities across cultures and languages in terms of the ability to recognize emotion from prosody [28]. Such invariance may even be contributed by the biological depth of prosody in potentially having evolutionary homologies among primates [30], as demonstrated by comparative studies showing parallels in the ways in which humans and animals shape spectro-temporal features of their vocalizations to express emotion [30–32]. To further understand the biological significance of prosodic differences (specifically with regard to rhythm) in ASD, future studies should examine the structural and functional neurophysiology associated with the production of prosody in speakers of multiple languages. For instance, functional neuroimaging studies have found increased right inferior frontal gyrus [76] and right caudate activation in individuals with ASD when performing prosody comprehension tasks [77], which were interpreted as evidence of more effortful processing needed to interpret prosodic cues. Such findings would be particularly compelling if replicated in individuals with ASD who are speakers of languages other than English.

In contrast with strong performance for classifications using rhythm-relevant acoustic features in both languages in Model 1, classifications using intonation-relevant acoustic features (timing-varying f_0 values) were significant for English, but not Cantonese. Unlike non-tone languages where pitch serves linguistic and para-linguistic functions at word, phrase, and utterance levels, in tone languages such as Cantonese, pitch is also used to convey meaning at the syllabic and lexical levels. Therefore, differences in intonation may manifest differently across languages due to cross-linguistic variability (which is represented in our intonation-relevant features, as suggested by the supplementary analysis presented in §3 in [S1 File](#)), into f_0 contour patterns that were identifiable by the ML algorithm for English but not for Cantonese. Contrasting prior findings showing common cross-linguistic pitch characteristics in ASD among typologically and prosodically related languages (i.e., English and Danish [41], Cantonese and Mandarin [36]), the cross-linguistic differences between English and Cantonese may have contributed to the lack of significant classifications using intonation-relevant features in Model 2, where no common feature patterns consistently associated with ASD across the two languages combined could be identified.

One consideration in cross-linguistic variability observed in ASD intonation is whether intonation differences in ASD are apparent only in speakers of English (or its typologically related languages [41]). Indeed, previous acoustic analyses only reported utterance-level f_0 differences across ASD and TD groups in English [17–19, 21, 22] but not Cantonese [35], consistent with our classification patterns using f_0 -based features across Models 1 and 2. It is possible that the prolific usage of linguistic pitch in tone languages provides a compensatory effect ameliorating intonational differences in ASD. In the perception domain, pitch processing differences found in tone language-speaking children with ASD [78–81] were surprisingly not evident in their adult peers [82], potentially due to a longer exposure to the native tone language. This possibility, i.e., that extensive pitch experience may ameliorate intonational differences in ASD, highlights pitch and intonation as a fruitful target for speech interventions in ASD for those who do not speak a tone language, where targeting this potentially more malleable factor could lead to therapeutic gains and help to advance more global speech and language characteristics. From a genetic perspective, in contrast to people of European descent, most Han Chinese people are carriers of the T allele of the ASPM gene that favors the ability to process linguistic pitch patterns [83, 84]. This may implicate a role of genetic factors in contributing to cross-linguistic differences of ASD phenotypes, such as by potentially ameliorating intonational differences in our participants from our Cantonese ASD group, all of whom

reported to be of Han Chinese descent. Future studies might examine the *ASPM* gene as a potential genetic marker for prosodic differences in ASD, shedding light onto the etiologies of communications disorders [85] that contribute to distinct cross-linguistic patterns of intonation in ASD.

Limitations

Several potential limitations should be considered in interpreting findings. First, there was a wide age range across our Cantonese and English groups and a small number of females included in the current study. With exhaustive efforts made to avoid model overfitting (i.e., the choice of linear SVM, repeated cross-validation, data reduction, and hyperparameter tuning in a nested fashion), robust classification results were found, potentially representing aspects of prosodic profiles of ASD which are relatively age and gender invariant. Moreover, neither age nor gender appeared to influence classification results, as per our post-hoc analysis (§2 in [S1 File](#)). Nevertheless, future studies with larger samples are needed to confirm whether findings reported here extend to both males and females and different age ranges where prosodic abilities may vary. Relatedly, future studies should employ more comprehensive assessments of language abilities for group matching, since baseline language ability may covary with age and autism severity [86, 87], influencing prosody and pitch in particular [88]. Finally, it will be important for future work to examine languages from typological classes beyond English and Cantonese. Such work will optimally include much larger, well matched corpora of multi-linguistic narrative samples from individuals with ASD, with potential to shed crucial insight into whether the rhythmic commonality identified here may represent features of ASD that are expressed across languages, despite cross-linguistic variations in prosodic properties.

Conclusion

Using brief utterance samples from a structured narrative task, ML models applied in this study were able to robustly classify individuals with ASD from those with TD (reflected by median AUC values ~ 0.9), based on rhythm-relevant features in both English- and Cantonese-speaking populations. The success of our ML approach yields implications for its future clinical utility, such as developing automatic detection of ASD to augment diagnosis [44] for personalizing therapy and training regimens [89]. Although further study with larger samples, and with additional language comparisons, is needed to support adaptation of the current models into clinical use, the current findings highlight the potential of ML-based methods to objectively identify other speech language-relevant differences in ASD (and potentially other neurodevelopmental disabilities impacting speech and language) using acoustic data [49]. Further, speech samples were relatively convenient and efficient to obtain, rendering these methods relatively feasible and potentially high yield for clinical application, and highlighting the promise of optimizing ML-based diagnostic models that make use of speech acoustic data as clinical tools. Future studies might also apply ML-based approaches to stratify populations for target in intervention and biological studies alike.

Together, findings from this study provide some of the first evidence for cross-linguistic commonalities in rhythmic characteristics in ASD across two typologically and prosodically distinct languages, while suggesting intonational characteristics in ASD are different across the two languages. Future studies examining languages rhythmically similar to English (e.g., German and Dutch), as well as those distant from both English and Cantonese (e.g., Japanese), could provide additional crucial insights into establishing prosodic profiles of ASD (e.g., rhythm) which could potentially be invariant across languages, while highlighting those (e.g., intonation) potentially more malleable to be shaped by language-specific linguistic properties.

Supporting information

S1 File. Supplementary methods, post-hoc analysis, and supplementary machine learning analysis.
(PDF)

Acknowledgments

We would like to thank families in the U.S. and Hong Kong who participated in this research study.

Author Contributions

Conceptualization: Joseph C. Y. Lau, Shivani Patel, Kritika Nayar, Patrick C. M. Wong, Molly Losh.

Data curation: Joseph C. Y. Lau, Shivani Patel, Xin Kang, Kritika Nayar, Gary E. Martin, Jason Choy.

Formal analysis: Joseph C. Y. Lau, Shivani Patel, Xin Kang, Kritika Nayar, Gary E. Martin, Jason Choy.

Funding acquisition: Patrick C. M. Wong, Molly Losh.

Investigation: Shivani Patel, Xin Kang, Gary E. Martin, Jason Choy, Molly Losh.

Methodology: Joseph C. Y. Lau, Gary E. Martin, Patrick C. M. Wong.

Project administration: Shivani Patel, Xin Kang, Patrick C. M. Wong, Molly Losh.

Resources: Joseph C. Y. Lau, Patrick C. M. Wong, Molly Losh.

Software: Joseph C. Y. Lau.

Supervision: Gary E. Martin, Patrick C. M. Wong, Molly Losh.

Validation: Molly Losh.

Visualization: Joseph C. Y. Lau.

Writing – original draft: Joseph C. Y. Lau, Molly Losh.

Writing – review & editing: Joseph C. Y. Lau, Shivani Patel, Xin Kang, Kritika Nayar, Jason Choy, Patrick C. M. Wong, Molly Losh.

References

1. McCann J, Peppé S. Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*. 2003; 38(4):325–350. <https://doi.org/10.1080/1368282031000154204> PMID: 14578051
2. Asghari SZ, Farashi S, Bashirian S, Jenabi E. Distinctive prosodic features of people with autism spectrum disorder: a systematic review and meta-analysis study. *Scientific reports*. 2021; 11(1):1–17. <https://doi.org/10.1038/s41598-021-02487-6> PMID: 34845298
3. Nakai Y, Takashima R, Takiguchi T, Takada S. Speech intonation in children with autism spectrum disorder. *Brain and Development*. 2014; 36(6):516–522. <https://doi.org/10.1016/j.braindev.2013.07.006> PMID: 23973369
4. Paul R, Augustyn A, Klin A, Volkmar FR. Perception and production of prosody by speakers with autism spectrum disorders. *Journal of autism and developmental disorders*. 2005; 35(2):205–220. <https://doi.org/10.1007/s10803-004-1999-1> PMID: 15909407

5. Dahlgren S, Sandberg AD, Strömbergsson S, Wenhov L, Råstam M, Nettelbladt U. Prosodic traits in speech produced by children with autism spectrum disorders—Perceptual and acoustic measurements. *Autism & Developmental Language Impairments*. 2018; 3:2396941518764527.
6. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*. 2017; 10(3):384–407. <https://doi.org/10.1002/aur.1678> PMID: 27501063
7. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. 5th ed. Washington, DC: Autor; 2013.
8. Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A. Autism and pervasive developmental disorders. *Journal of child psychology and psychiatry*. 2004; 45(1):135–170. <https://doi.org/10.1046/j.0021-9630.2003.00317.x> PMID: 14959806
9. Mitchell S, Brian J, Zwaigenbaum L, Roberts W, Szatmari P, Smith I, et al. Early language and communication development of infants later diagnosed with autism spectrum disorder. *Journal of Developmental & Behavioral Pediatrics*. 2006; 27(2):S69–S78. <https://doi.org/10.1097/00004703-200604002-00004> PMID: 16685188
10. Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*. 2006; 38(10):1776–1791. <https://doi.org/10.1016/j.pragma.2005.07.004>
11. Cutler A, Isard SD. The production of prosody. In: Butterworth B, editor. *Language production*. London: Academic Press; 1980. p. 245–269.
12. Nooteboom S. The prosody of speech: melody and rhythm. *The handbook of phonetic sciences*. 1997; 5:640–673.
13. Fox A. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford: Oxford University Press; 2000.
14. Patel SP, Nayar K, Martin GE, Franich K, Crawford S, Diehl JJ, et al. An Acoustic Characterization of Prosodic Differences in Autism Spectrum Disorder and First-Degree Relatives. *Journal of Autism and Developmental Disorders*. 2020; p. 1–14. <https://doi.org/10.1007/s10803-020-04392-9> PMID: 32056118
15. Beckman ME. Evidence for speech rhythms across languages. *Speech perception, production and linguistic structure*. 1992; p. 457–463.
16. Kanner L. Autistic disturbances of affective contact. *Nervous child*. 1943; 2(3):217–250.
17. Edelson L, Grossman R, Tager-Flusberg H. Emotional prosody in children and adolescents with autism. In: Poster session presented at the annual international meeting for Autism Research, Seattle, WA; 2007.
18. Diehl JJ, Watson D, Bennetto L, McDonough J, Gunlogson C. An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*. 2009; 30(3):385. <https://doi.org/10.1017/S0142716409090201>
19. Fosnot SM, Jun S. Prosodic characteristics in children with stuttering or autism during reading and imitation. In: *Proceedings of the 14th international congress of phonetic sciences*; 1999. p. 1925–1928.
20. Hubbard DJ, Faso DJ, Assmann PF, Sasson NJ. Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research*. 2017; 10(12):1991–2001. <https://doi.org/10.1002/aur.1847> PMID: 28815940
21. Nadig A, Shaw H. Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*. 2012; 42(4):499–511. <https://doi.org/10.1007/s10803-011-1264-3> PMID: 21528425
22. Paul R, Bianchi N, Augustyn A, Klin A, Volkmar FR. Production of syllable stress in speakers with autism spectrum disorders. *Research in autism spectrum disorders*. 2008; 2(1):110–124. <https://doi.org/10.1016/j.rasd.2007.04.001> PMID: 19337577
23. Diehl JJ, Paul R. The assessment and treatment of prosodic disorders and neurological theories of prosody. *International journal of speech-language pathology*. 2009; 11(4):287–292. <https://doi.org/10.1080/17549500902971887> PMID: 20852744
24. Mesibov GB. Treatment issues with high-functioning adolescents and adults with autism. In: *High-functioning individuals with autism*. New York: Springer; 1992. p. 143–155.
25. Van Bourgondien ME, Woods AV. Vocational possibilities for high-functioning adults with autism. In: *High-functioning individuals with autism*. New York: Springer; 1992. p. 227–239.
26. Seddoh SA. How discrete or independent are “affective prosody” and “linguistic prosody”? *Aphasiology*. 2002; 16(7):683–692. <https://doi.org/10.1080/02687030143000861>

27. Scherer KR. Vocal affect expression: A review and a model for future research. *Psychological bulletin*. 1986; 99(2):143. <https://doi.org/10.1037/0033-2909.99.2.143> PMID: 3515381
28. Cowen AS, Laukka P, Effenbein HA, Liu R, Keltner D. The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature human behaviour*. 2019; 3(4):369–382. <https://doi.org/10.1038/s41562-019-0533-6> PMID: 30971794
29. Paulmann S, Uskul AK. Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & emotion*. 2014; 28(2):230–244. <https://doi.org/10.1080/02699931.2013.812033> PMID: 23862740
30. Filippi P. Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Frontiers in Psychology*. 2016; 7:1393. <https://doi.org/10.3389/fpsyg.2016.01393> PMID: 27733835
31. Filippi P, Congdon JV, Hoang J, Bowling DL, Reber SA, Pašukonis A, et al. Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*. 2017; 284(1859):20170990. <https://doi.org/10.1098/rspb.2017.0990> PMID: 28747478
32. Parr LA, Waller BM, Vick SJ. New developments in understanding emotional facial signals in chimpanzees. *Current Directions in Psychological Science*. 2007; 16(3):117–122. <https://doi.org/10.1111/j.1467-8721.2007.00487.x> PMID: 20182651
33. Abercrombie D. *Elements of general phonetics*. Chicago: Aldine Pub. Company; 1967.
34. Thongseitratch T, Chuthapisith J, Roengpitya R. An acoustic study of the five Thai tones produced by ASD and TD children. In: *The Proceedings of the 15th Australasian International Speech Science & Technology Conference SST 2014*; 2014. p. 106–109.
35. Chan KK, To CK. Do individuals with high-functioning autism who speak a tone language show intonation deficits? *Journal of autism and developmental disorders*. 2016; 46(5):1784–1792. <https://doi.org/10.1007/s10803-016-2709-5> PMID: 26825662
36. Chen F, Cheung CCH, Peng G. Linguistic tone and non-linguistic pitch imitation in children with autism spectrum disorders: A cross-linguistic investigation. *Journal of Autism and Developmental Disorders*. 2021; p. 1–19. PMID: 34109462
37. Mok P. On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*. 2009; 2:148–154.
38. Effenbein HA, Ambady N. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*. 2002; 128(2):203. <https://doi.org/10.1037/0033-2909.128.2.203> PMID: 11931516
39. Russell JA. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*. 1994; 115(1):102. <https://doi.org/10.1037/0033-2909.115.1.102> PMID: 8202574
40. Thompson WF, Balkwill LL. Decoding speech prosody in five languages. *Semiotica*. 2006; 2006(158):407–424. <https://doi.org/10.1515/SEM.2006.017>
41. Fusaroli R, Grossman R, Bilenberg N, Cantio C, Jepsen JRM, Weed E. Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Research*. 2021;. PMID: 34957701
42. Oller DK, Niyogi P, Gray S, Richards JA, Gilkerson J, Xu D, et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*. 2010; 107(30):13354–13359. <https://doi.org/10.1073/pnas.1003882107> PMID: 20643944
43. Santos JF, Brosh N, Falk TH, Zwaigenbaum L, Bryson SE, Roberts W, et al. Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE; 2013. p. 7567–7571.
44. Nakai Y, Takiguchi T, Matsui G, Yamaoka N, Takada S. Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists. *Perceptual and motor skills*. 2017; 124(5):961–973. <https://doi.org/10.1177/0031512517716855> PMID: 28649923
45. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*. 2017; 81:181–187. <https://doi.org/10.1016/j.neubiorev.2017.02.011> PMID: 28212857
46. Tilsen S, Johnson K. Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*. 2008; 124(2):EL34–EL39. <https://doi.org/10.1121/1.2947626> PMID: 18681499

47. Tilsen S, Arvaniti A. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*. 2013; 134(1):628–639. <https://doi.org/10.1121/1.4807565> PMID: 23862837
48. Hyde KK, Novack MN, LaHaye N, Parlett-Pelleriti C, Anden R, Dixon DR, et al. Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*. 2019; 6(2):128–146. <https://doi.org/10.1007/s40489-019-00158-x>
49. Lauraitis A, Maskeliūnas R, Damaševičius R, Krilavičius T. Detection of Speech Impairments Using Cepstrum, Auditory Spectrogram and Wavelet Time Scattering Domain Features. *IEEE Access*. 2020; <https://doi.org/10.1109/ACCESS.2020.2995737>
50. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020; 5(1):96–116. <https://doi.org/10.1002/liv.2.354> PMID: 32128436
51. Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. *Current Alzheimer Research*. 2020; 17(1):60. <https://doi.org/10.2174/1567205017666200213094513> PMID: 32053074
52. Hus V, Gotham K, Lord C. Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of autism and developmental disorders*. 2014; 44(10):2400–2412. <https://doi.org/10.1007/s10803-012-1719-1> PMID: 23143131
53. Mayer M. *Frog, where are you?* New York: Dial Press; 1969.
54. Capps L, Losh M, Thurber C. “The frog ate the bug and made his mouth sad”: Narrative competence in children with autism. *Journal of abnormal child psychology*. 2000; 28(2):193–204. <https://doi.org/10.1023/A:1005126915631> PMID: 10834770
55. Diehl JJ, Bennetto L, Young EC. Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of abnormal child psychology*. 2006; 34(1):83–98. <https://doi.org/10.1007/s10802-005-9003-x> PMID: 16485176
56. Losh M, Capps L. Narrative ability in high-functioning children with autism or Asperger’s syndrome. *Journal of autism and developmental disorders*. 2003; 33(3):239–251. <https://doi.org/10.1023/A:1024446215446> PMID: 12908827
57. Reilly J, Losh M, Bellugi U, Wulfeck B. “Frog, where are you?” Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and language*. 2004; 88(2):229–247. [https://doi.org/10.1016/S0093-934X\(03\)00101-9](https://doi.org/10.1016/S0093-934X(03)00101-9) PMID: 14965544
58. Tager-Flusberg H, Sullivan K. Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*. 1995; 16(3):241–256. <https://doi.org/10.1017/S0142716400007281>
59. Berman RA, Slobin DI, editors. *Relating events in narrative: A crosslinguistic developmental study*. Mahwah: Lawrence Erlbaum Associates, Inc.; 1994.
60. Himmelman NP, Sandler M, Strunk J, Unterladstetter V. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology*. 2018; 35(2):207–245. <https://doi.org/10.1017/S0952675718000039>
61. Ladd DR. *Intonational phonology*. Cambridge: Cambridge University Press; 2008.
62. Selkirk E. 14 The Syntax-Phonology Interface. *The handbook of phonological theory*. 2011; p. 435. <https://doi.org/10.1002/9781444343069.ch14>
63. Silverman K, Beckman M, Pitrelli J, Ostendorf M, Wightman C, Price P, et al. ToBI: A standard for labeling English prosody. In: *Second international conference on spoken language processing*; 1992.
64. Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*. 2020; p. 1–13. PMID: 32376899
65. Hickok G, Poeppel D. The cortical organization of speech processing. *Nature Reviews Neuroscience*. 2007; 8:393–402. <https://doi.org/10.1038/nrn2113> PMID: 17431404
66. Wong PC. Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*. 2002; 59(2):83–95. [https://doi.org/10.1016/S0361-9230\(02\)00860-2](https://doi.org/10.1016/S0361-9230(02)00860-2) PMID: 12379438
67. Liu C, Cheng Y. An application of the support vector machine for attribute-by-attribute classification in cognitive diagnosis. *Applied psychological measurement*. 2018; 42(1):58–72. <https://doi.org/10.1177/0146621617712246> PMID: 29881112
68. Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*. 2012; 36(4):1140–1152. <https://doi.org/10.1016/j.neubiorev.2012.01.004> PMID: 22305994

69. Zhang X, Wong WH. Recursive sample classification and gene selection based on SVM: method and software description. Biostatistics Dpt Tech Report, Harvard School of Public Health. 2001.
70. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS one*. 2019; 14(11):e0224365. <https://doi.org/10.1371/journal.pone.0224365> PMID: 31697686
71. Baltaxe CA, Simmons JQ. Prosodic development in normal and autistic children. In: Communication problems in autism. New York: Springer; 1985. p. 95–125.
72. Baron-Cohen S, Staunton R. Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First language*. 1994; 14(42-43):241–248. <https://doi.org/10.1177/014272379401404216>
73. Fay WH. On the basis of autistic echolalia. *Journal of Communication Disorders*. 1969; 2(1):38–47. [https://doi.org/10.1016/0021-9924\(69\)90053-7](https://doi.org/10.1016/0021-9924(69)90053-7)
74. Pronovost W, Wakstein MP, Wakstein DJ. A longitudinal study of the speech behavior and language comprehension of fourteen children diagnosed atypical or autistic. *Exceptional children*. 1966; 33(1):19–26. <https://doi.org/10.1177/001440296603300104> PMID: 5981255
75. Shriberg LD, Paul R, McSweeney JL, Klin A, Cohen DJ, Volkmar FR. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*. 2001;. [https://doi.org/10.1044/1092-4388\(2001/087\)](https://doi.org/10.1044/1092-4388(2001/087)) PMID: 11708530
76. Wang AT, Lee SS, Sigman M, Dapretto M. Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain*. 2006; 129(4):932–943. <https://doi.org/10.1093/brain/awl032> PMID: 16481375
77. Gebauer L, Skewes J, Hørlyck L, Vuust P. Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical*. 2014; 6:370–378. <https://doi.org/10.1016/j.nicl.2014.08.025> PMID: 25379450
78. Lau JC, To CK, Kwan JS, Kang X, Losh M, Wong PC. Lifelong tone language experience does not eliminate deficits in neural encoding of pitch in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2020; p. 1–20.
79. Wang X, Wang S, Fan Y, Huang D, Zhang Y. Speech-specific categorical perception deficit in autism: an event-related potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*. 2017; 7. <https://doi.org/10.1038/srep43254> PMID: 28225070
80. Yu L, Fan Y, Deng Z, Huang D, Wang S, Zhang Y. Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *Journal of autism and developmental disorders*. 2015; 45(11):3656–3667. <https://doi.org/10.1007/s10803-015-2510-x> PMID: 26111738
81. Zhang J, Meng Y, Wu C, Xiang YT, Yuan Z. Non-speech and speech pitch perception among Cantonese-speaking children with autism spectrum disorder: An ERP study. *Neuroscience letters*. 2019; 703:205–212. <https://doi.org/10.1016/j.neulet.2019.03.021> PMID: 30890474
82. Cheng ST, Lam GY, To CK. Pitch Perception in Tone Language-Speaking Adults With and Without Autism Spectrum Disorders. *i-Perception*. 2017; 8(3):2041669517711200. <https://doi.org/10.1177/2041669517711200> PMID: 28616150
83. Wong PC, Chandrasekaran B, Zheng J. The derived allele of ASPM is associated with lexical tone perception. *PLoS one*. 2012; 7(4):e34243. <https://doi.org/10.1371/journal.pone.0034243> PMID: 22529908
84. Wong PC, Kang X, Wong KH, So HC, Choy KW, Geng X. ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*. 2020; 6(22):eaba5090. <https://doi.org/10.1126/sciadv.aba5090> PMID: 32537487
85. Wong PC, Perrachione TK, Gunasekera G, Chandrasekaran B. Communication disorders in speakers of tone languages: etiological bases and clinical considerations. In: *Seminars in speech and language*. vol. 30. © Thieme Medical Publishers; 2009. p. 162–173.
86. Ravid D. Emergence of linguistic complexity in later language development: Evidence from expository text construction. In: *Perspectives on language and language development*. Springer; 2005. p. 337–355.
87. Lust B, Foley C, Dye CD. The first language acquisition of complex sentences. *Cambridge Handbook of Child Language*. 2009.
88. Wagner M, Watson DG. Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*. 2010; 25(7-9):905–945. <https://doi.org/10.1080/01690961003589492> PMID: 22096264
89. Wong PC, Vuong LC, Liu K. Personalized learning: From neurogenetics of behaviors to designing optimal language training. *Neuropsychologia*. 2017; 98:192–200. <https://doi.org/10.1016/j.neuropsychologia.2016.10.002> PMID: 27720749