10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Assessing the Intonation Style of Speakers with Autism Spectrum Disorder

Simon Wehrle[1], Francesco Cangemi[1], Harriet Hanekamp[1], Kai Vogeley[2,3] & Martine Grice[1]

[1] IfL-Phonetik, University of Cologne, Germany; [2] Research Centre Juelich, Institute of Neuroscience and Medicine (INM3), Germany;
[3] Dept. Psychiatry, Faculty of Medicine and University Hospital Cologne, Germany;
{simon.wehrle,fcangemi,hhaneka1,martine.grice}@uni-koeln.de,
kai.vogeley@uk-koeln.de

Abstract

Speakers with Autism Spectrum Disorder (ASD) have been claimed, and are generally assumed, to produce atypical intonation. Previous research, however, is very limited and findings are contradictory, with claims ranging from "robotic" to "singsongy" intonation styles in ASD. We employ a novel method to assess intonation styles in a large corpus of semi-spontaneous speech by German adults with ASD and matched controls. We show that the ASD group has a more melodic or singsongy intonation style as a whole, but that this effect is subject to a high degree of individual variation and that the difference between groups is greater for male than for female speakers. We compare the method of analysis proposed here with the more traditional measures of pitch range and mean f0, showing that the novel approach is able to capture the difference between intonation styles more clearly.

Index Terms: intonation, autism, pitch range, Map Task

1. Introduction

Intonation styles have been used in linguistics to characterise a number of language varieties and groups of speakers [1-4]. For the group of individuals with Autism Spectrum Disorder (ASD), intonation is generally assumed to be "odd" or "atypical", but there is no consensus on what this actually means. The very limited number of studies that have attempted to identify aspects of intonation characteristic for speakers with ASD to date have yielded inconclusive and contradictory results. The claims reach from "singsongy" intonation (with e.g. an increased pitch range) on the one hand [5,32] to the very opposite notion of monotonous or "robotic"-sounding speech [6,7]. Both claims are consistent with [8], who found both very narrow and very wide pitch ranges in children with ASD.

There are at least three possible reasons for this uncertainty regarding the nature of intonation styles in ASD. First, the speech data used in previous studies was mostly elicited through reading tasks or structured interviews, methods that are not conducive to natural intonation. Second, we have to consider the fact that speakers with ASD constitute a very heterogeneous group characterised by a high degree of individual variation. If speaker-specific behaviour is not appropriately taken into account, as has all too often been the case in previous research, averaged values alone cannot be expected to paint a realistic picture of either the behaviour of the group as a whole or of any of the individuals within it [9]. Third, where characterisation of intonation styles has gone beyond simple subjective judgement in the past, the acoustic measures used (mostly pitch range and long-term distributional measures of f0, see [4,10]) may in themselves not be sufficient to distinguish even extreme cases such as stereotypically robotic or singsongy speech. In [11] we have examined this issue in detail and proposed a new method for capturing intonation styles by focussing on the time-varying dynamics of pitch contours along two dimensions, which we term Wiggliness and Spaciousness. For the present contribution, we have modified, refined and automated this method and use it in an in-depth analysis of semi-spontaneous speech by German adults with and without a diagnosis of ASD. We thereby avoid a number of shortcomings affecting previous studies and hope to shed some light onto the true nature of intonation styles in speakers with ASD.

2. Method

2.1. Material

We recorded Map Tasks to elicit semi-spontaneous speech in the form of task-oriented dialogues [12,13]. In the Map Task paradigm, participants are recorded in dyads upon being presented with a set of simple maps. Only one of the two participants has a route printed on their map. The experimental task is for the first speaker to describe their route to the interlocutor as precisely as possible, such that the latter can reproduce the instruction giver's route on their own map. In the process, participants do not have any visual contact and must solve the task by means of verbal communication alone. All dyads completed two Map Tasks, switching the roles of instruction giver and instruction follower after completion of the first task. Map Task conversations were recorded in a sound-proof booth at the Department of Phonetics of the University of Cologne. We used two head-mounted microphones (AKG C420L) connected through an audio-interface (PreSonus AudioBox 22VSL) to a PC running *Adobe Audition* with a sample rate of 44100 Hz (16 bit).

Recordings were transcribed orthographically and divided into interpausal units (IPUs) with a minimum pause length of 200ms for each speaker. All IPUs contained within the first five minutes of each participant's turn as instruction giver were automatically extracted and then manually corrected for pitch errors and smoothed [14]. For some speakers, Map Tasks were completed in under five minutes, in which case we simply extracted and analysed the entirety of the speech material. We excluded all IPUs with a duration of less than 1 second from further analysis as such excerpts do not contain enough speech

 10.21437/SpeechProsody.2020-165

material for a dynamic characterisation of intonation styles. A very large proportion of these short IPUs consist of backchannels (listeners signals such as “mmhm” or “yeah”) which are being analysed separately for a parallel project. After exclusion of very short IPUs and 9 items that were found to be based on unreliable pitch extraction, 1483 IPUs (with a mean duration of 2.81 seconds) remained for analysis.

2.2. Participants

28 native speakers of German (14 ASD, 14 controls) took part in the experiment. Participants from the ASD group had all been diagnosed with ICD-10: F84.5 (Asperger’s Syndrome [15]) and were recruited in the Autism Outpatient Clinic at the Department of Psychiatry at the University of Cologne (Germany). Participants from the control group (CTR) were recruited from the general population and were paid 10 euros each for participation. All participants completed the German version of the Autism-Spectrum Quotient (AQ) questionnaire [16]. AQ scores range from 0 to 50, with higher scores indicating more autistic traits. All participants also completed the *Wortschatztest WST* [17], a standardized German vocabulary test that can serve as a measure of both verbal and general IQ [18].
Although participants were matched as closely as possible for gender, age and IQ, gender ratios differed very slightly between groups and participants from the ASD group were on average slightly older and had a slightly higher verbal IQ than participants from the CTR group (see Table 1). Crucially, however, participants from the ASD group had a far higher average AQ score than controls and there was no overlap whatsoever in AQ scores between the two groups.

Table 1: *Subject information by group. Mean values (standard deviations in brackets).*

	Gender	Age	Verbal IQ	AQ
ASD	10 m; 4 f	42.5 (7.8)	116.9 (12.6)	41.4 (3.3)
CTR	11 m; 3 f	37.3 (8)	106.4 (6.1)	15.7 (4.3)

2.3. Analysis

For assessment of intonation styles, we employed a novel approach, focussing on the time-varying dynamics of pitch contours, represented by two parameters: Wiggliness and Spaciousness (see [11] for rationale and detailed description).
Wiggliness is operationalised as the amount of times an f0 contour “changes direction” in a given amount of time, i.e. as slope changes per second. To obtain the measure of Wiggliness, smoothed and corrected pitch contours were automatically stylised to a resolution of 2 semitones in Praat [19]. An automatic procedure then computed the number of rises and falls within the stylised pitch object and divided this number by the duration of the pitch object. Wiggliness values range from 0.6 to 8.07 in our dataset (with a global mean of 3.22 (SD = 1.02)).
Spaciousness is operationalised as the extent of the slopes of the individual rises and falls within a given IPU, i.e. the maximum f0 excursions. The Spaciousness measure (in semitones) was automatically computed as the average between the absolute values of the two largest excursions within an IPU and ranges from 0.28 ST to 14.75 ST in our data (with a global mean of 6.48 Hz (SD = 2.45)). Semitones (with a reference value of 1 Hz) were chosen as the unit of measurement rather than Hertz (as in [11]) for being perceptually more valid [31] and for facilitating comparison between male and female speakers.
Based on [11] we can assume that the more wiggly and spacious a pitch contour is, the more “singsongy” or melodic the intonation style it represents; the less wiggly and spacious the contour, the more “robotic” or monotonous the intonation style it represents. Wiggliness and Spaciousness are correlated, but partly independent measures, and will be plotted and reported together in the following analysis. For comparison with measures used in previous studies, we also report values of pitch range and mean f0. Pitch range was calculated as the difference between the maximum and minimum of (corrected) f0 values in each IPU. Mean f0 was calculated as the average of all f0 values across IPUs.
Since this study is exploratory rather than confirmatory in nature, we will not be reporting inferential statistics in a null hypothesis significance testing framework. In the absence of reliable previous research and relevant pilot data, it would be misleading to posit distinct hypotheses to be tested using methods relying on statistical significance [20-24]. Such an approach would inherently involve an increased risk of irresponsibly disseminating spurious results based on Type I errors. We will instead rely on a comprehensive and transparent analysis of the data at hand, combined with an emphasis on descriptive and visual clarity in order to sufficiently convey all relevant findings. All data and scripts are available for download and inspection at https://osf.io/mt8fn/.

3. Results

Figure 1 shows mean Wiggliness and Spaciousness values by group. The ASD group has higher Wiggliness and higher Spaciousness overall than the CTR group.

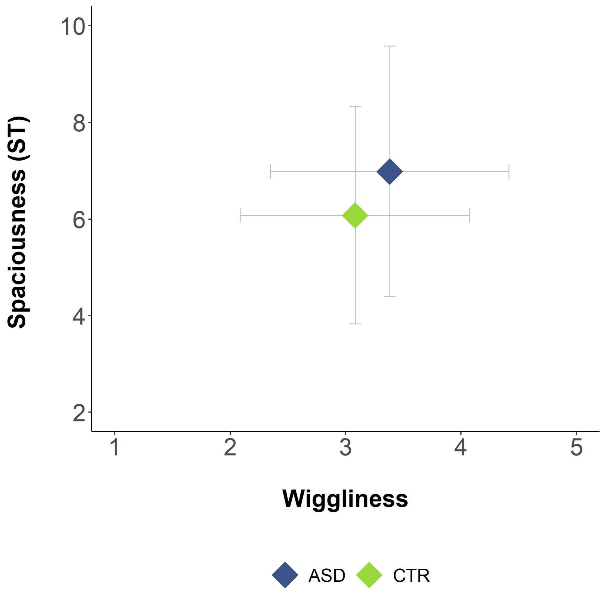

Figure 1: *Mean Spaciousness (in ST, on the y-axis) and Wiggliness (on the x-axis) by group. Error bars represent one standard deviation from the mean.*

This between-group difference holds true across genders, but the difference between male speakers from either group is greater than that between female speakers (see Table 2).

Table 2: *Mean Wiggliness and Spaciousness by gender and group (standard deviations in brackets). Higher respective values in bold.*

		Wiggliness	Spaciousness
Male	*ASD*	**3.45 (1.07)**	**7.03 ST (2.48)**
	CTR	3.06 (0.98)	6.04 ST (2.26)
Female	*ASD*	**3.22 (0.93)**	**6.87 ST (2.86)**
	CTR	3.14 (1.03)	6.18 ST (2.2)

Figure 2 allows for a closer look at the data in presenting speaker-specific values. Although the overall pattern of higher Wiggliness and higher Spaciousness in the ASD group remains clear to see, there is also evidence of a considerable amount of speaker-specific variation and a high degree of overlap between the two groups.

The global impression of more singsongy and less robotic speech in ASD is corroborated by the fact that the 5 highest mean Spaciousness values and 5 out of the 6 highest mean Wiggliness values were produced by speakers from the ASD group, whilst the 5 lowest Spaciousness values and the 4 lowest Wiggliness values were produced by speakers from the CTR group.

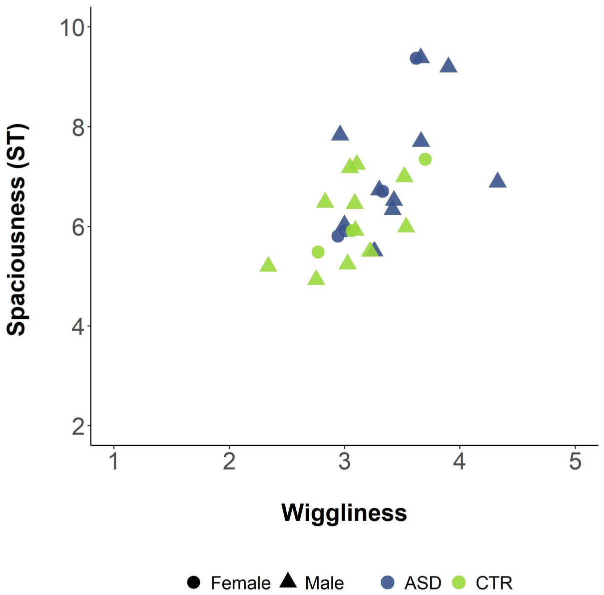

Figure 2: *Mean Spaciousness (in ST, on the y-axis) and Wiggliness values (on the x-axis) by speaker, group and gender.*

For comparison with previous studies, we calculated the global measures of pitch range and mean f0 for all speakers.

Pitch range is higher for ASD than for CTR speakers, and the difference between groups is greater for male than for female speakers (see Table 3), in line with the Wiggliness and Spaciousness results discussed above. The measures of pitch range and Spaciousness are clearly correlated, but neither identical nor equivalent. Notably, Spaciousness allows for a better separation of speakers by group rather than only by gender (see Fig. 3).

Table 3: *Pitch range and mean f0 in Hz, by gender and group (standard deviations in brackets). Higher respective values in bold.*

		Pitch range	Mean f0
Male	*ASD*	**91 Hz (46)**	**136 Hz (32)**
	CTR	70 Hz (38)	128 Hz (32)
Female	*ASD*	**143 Hz (60)**	226 Hz (41)
	CTR	134 Hz (56)	**237 Hz (50)**

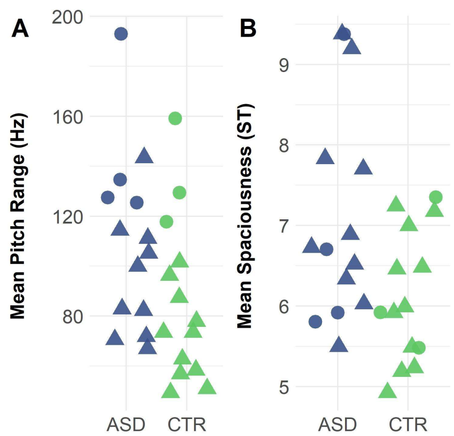

Figure 3: *Mean pitch range (y-axis in Panel A) and Spaciousness (y-axis in Panel B) by speaker, group and gender. Dots represent females, triangles males.*

Mean f0 values are highly similar between groups and thereby do not reveal the patterns found for the other measures discussed above (see Table 3). A speaker-specific analysis based on meanf0 is inconclusive beyond the fact that most of the lowest values are produced by males from the CTR group (see Fig. 4).

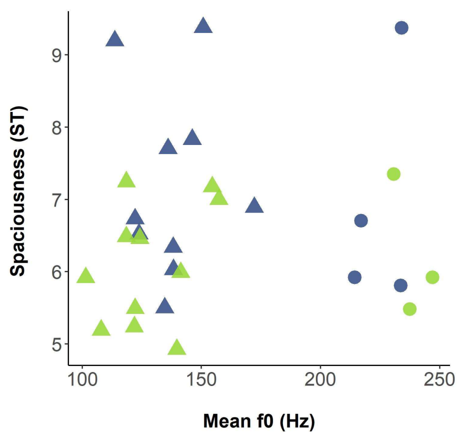

Figure 4: *Mean Spaciousness (y-axis) and mean f0 (x-axis) values by speaker, group and gender. Dots represent females, triangles males.*

4. Discussion

Our data show that, overall, German adults with ASD have a more melodic or “singsongy” and less monotonous or “robotic” intonation style than control speakers without a diagnosis of ASD, represented through higher values for both Wiggliness and Spaciousness (and a higher mean pitch range). Although both highly melodic and largely monotonous speech have been reported for ASD in the past, there seems to be a growing consensus that a more “singsongy” style with increased pitch range is closer to the true nature of speech in ASD (see [32] and references therein). The present contribution adds to this evidence.

A closer inspection of the data reveals that the difference between groups is slightly stronger for male speakers than for female speakers. As there is a male-to-female ratio of approximately 3:1 in our dataset, reflecting the real-life ratio of ASD diagnosis [25], the amount of data available for male vs. female participants is not comparable. To rule out that this could be an underlying reason for the diverging results between genders, we ran additional analyses by comparing all 7 females speakers to repeated iterations of random subsets of 7 male speakers in order to obtain gender ratios of 1:1. In all cases, the pattern that had also been observed for the entire dataset held true.

The fact that female speakers with ASD seem to diverge less from the intonation style of female controls than is the case for male speakers can be related to the observation that females with ASD are more likely than males to successfully mask their symptoms through the use of compensation strategies or “social camouflaging” [26-30]. Our results could reflect an instance of such social adaptation in the domain of speech prosody.

Beyond these group and gender differences, a high degree of individual variation can be observed in our dataset; there is no one distinct boundary that can be drawn between the CTR and the ASD group. We observe instead a considerable amount of overlap at the level of individual speakers. Thus, whilst the overall impression of a more singsongy and less robotic intonation style for ASD speakers is corroborated at the level of the individual, some speakers from the ASD group (8 out of 14) show behaviour that is well within the typical range of the CTR group. This level of analysis confirms the importance of individual specificity for speech research, all the more when it is concerned with atypical and heterogeneous groups of speakers, as is the case here.

We have shown that the newly proposed measures of Wiggliness and Spaciousness are related to but distinct from conventional measures like pitch range and mean f0. Based on our results, mean f0 does not appear to be a suitable metric for capturing intonation styles, unlike pitch range, which provided similar results to Wiggliness/Spaciousness. Moreover, as the multi-dimensional, dynamic approach to assessing intonation styles suggested here provides a clearer separation of speakers from the ASD and CTR groups than pitch range, it appears to be more suitable for uncovering the complex and subtle patterns of vocal behaviour which may fundamentally shape the impression of prosodic “oddness” in speakers with ASD.

Despite focusing on a rigorous and in-depth analysis, the results presented in this study may have been affected by a number of methodological issues. First, although task-oriented dialogue in the Map Task paradigm is a major improvement on read, scripted or formally constrained speech, it is still not entirely natural or spontaneous. Having to fulfil a task puts certain pressures and limitations on participants and their linguistic output; it may have affected speakers in the ASD group differently than those in the CTR group.

Second, the method used here to assess intonation styles presents a departure from traditional analyses and has not been rigorously tested on other datasets so far. Although the work in [11] and careful inspection of the data has reassured us that the two dimensions of Wiggliness and Spaciousness do indeed capture and distinguish “robotic” and “singsongy” intonation styles and what lies in-between, the precise measurements are still open to change and might be improved in future. For instance, it might be more suitable to use more than the two maximum excursions to calculate Spaciousness.

Third, comparison of the male and female groups suffers from the fact that more data is available for male than for female speakers, although additional analyses seemed to rule out that this a reason underlying the between-gender differences in our data. Future studies should nevertheless aim to also investigate groups with equal numbers of male and female speakers. More generally, a greater number of speakers would allow us to draw more reliable conclusions and corroborate any current working hypotheses as presented in this study.

5. Conclusion

We assessed the intonation style of German-speaking adults with and without ASD using a novel method of analysis. We showed that this new method is more appropriate than static, global measurements such as pitch range and mean f0 for revealing the differences between groups and individuals. Overall, the ASD group has a more singsongy and less monotonous intonation style than the CTR group and this between-group difference is more pronounced for male than for female speakers. However, differences between groups are subtle and only a speaker-specific analysis can do justice to the complex patterns of behavior characteristic of groups of individuals diagnosed with ASD.

6. Acknowledgements

We would like to thank three anonymous reviewers for their helpful comments. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 281511265 – SFB 1252 Prominence in Language) and a doctoral scholarship awarded to SW by the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes).

7. References

[1] L. Kuiper, “Variation and the norm: Parisian perceptions of regional French”, in *Handbook of Perceptual Dialectology Vol. I*, D. Preston, Ed. Amsterdam-Philadelphia: John Benjamins, 1999, pp. 243-262.

[2] A.-L. Coquillon and J. Durand, “Le français meridional: éléments de synthèse”, in *Les Variétés du Français Parlé dans l’Espace Francophone: Ressources pour l’Enseignement,* S. Detey, J. Durand, B. Laks & C. Lyche, Eds. Paris: Ophrys, 2010, pp. 185-197.

[3] M. Celce-Murcia, D. Brinton and J. Goodwin, *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press, 1996.

[4] C. Graham, “Fundamental frequency range in Japanese and English: the case of simultaneous bilinguals”, *Phonetica*, 71(4), pp. 271-295, 2014.

[5] J.Q. Simmons & C. Baltaxe, "Language patterns of adolescent autistics", *Journal of Autism and Childhood Schizophrenia*, 5(4), pp. 333-351, 1975.
[6] L. Kanner, "Autistic disturbances of affective contact", *Nervous Child*, 2(3), pp. 217-250, 1943.
[7] H. Green and Y. Tobin, "Prosodic analysis is difficult…but worth it: a study in high-functioning autism", *International Journal of Speech-Language Pathology,* 11(4), pp. 308-315, 2009.
[8] C. Baltaxe, "Use of contrastive stress in normal, aphasic and autistic children", *Journal of Speech, Language, and Hearing Research*, 27(1), 97-105, 1984.
[9] F. Cangemi, D. El Zarka, S. Wehrle, S. Baumann & M. Grice, "Speaker-specific intonational marking of narrow focus in Egyptian Arabic", *Proceedings of the 8th Speech Prosody Conference*, Boston, 2016, pp. 335-339, 2016.
[10] Mennen, F. Schaeffler and G. Docherty, "Cross-language differences in f0 range: a comparative study of English and German", in *Journal of the Acoustical Society of America*, 131(3), pp. 2249-2260, 2012.
[11] S. Wehrle, F. Cangemi, M. Krüger and M. Grice, "Somewhere over the spectrum: between singsongy and robotic intonation", in *Studi AISV 4: Speech in the Natural Context*, A. Vietti, L, Spreafico, D. Mereu and V. Galatà, Eds., 2018.
[12] G. Brown, A. Anderson, R. Shillcock & G. Yule, *Teaching Talk: Strategies for Production and Assessment*, Cambridge: Cambridge University Press, 1985.
[13] A.H. Anderson et al., "The HCRC Map Task corpus", *Language and Speech*, 34(4), 351-366, 1991.
[14] F. Cangemi, *mausmooth*. 2015.
[15] World Health Organization, The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization, 1992.
[16] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin and E. Clubley, "The autism spectrum quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians", *Journal of Autism and Developmental Disorders*, 31(1), pp. 5-17, 2001.
[17] K.-H. Schmidt and P. Metzler, *Wortschatztest: WST.* Weinheim: Beltz, 1992.
[18] W. Satzger, H. Fessmann, and R.R. Engel, „Liefern HAWIE-R, WST und MWT-B vergleichbare IQ-Werte?", *Zeitschrift fur Differentielle und Diagnostische Psychologie,* 23(2), pp. 159–170, 2002.
[19] P. Boersma & D. Weenink, Praat: Doing Phonetics by Computer, 2019.
[20] J. W. Tukey, "We need both exploratory and confirmatory", *The American Statistician*, 34(1), 23-25, 1980.
[21] V. Amrhein and S. Greenland, "Remove, rather than redefine, statistical significance", *Nature Human Behaviour*, 2(1), p. 4, 2018.
[22] V. Amrhein, S. Greenland and B. McShane, "Scientists rise up against statistical significance", *Nature,* 567, pp. 305– 307, 2019.
[23] T. B. Roettger, "Researcher degrees of freedom in phonetic research", *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(1), 2019.
[24] T.B. Roettger, B. Winter & H. Baayen, "Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility", *Journal of Phonetics*, 73, pp. 1-7, 2019.
[25] R. Loomes, L. Hull and W.P.L. Mandy, "What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis", *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(6), pp. 466-474, 2017.
[26] T.M. Krahn and A. Fenton, "The extreme male brain theory of autism and the potential adverse effects for boys and girls with autism", *Journal of Bioethical Inquiry,* 9(1), pp. 93-103, 2012.
[27] S. Begeer, D. Mandell, B. Wijnker-Holmes, S. Venderbosch, D. Rem, F. Stekelenburg, and H.M. Koot, "Sex differences in the timing of identification among children and adults with autism spectrum disorders", *Journal of Autism and Developmental Disorders*, 43(5), pp. 1151-1156, 2013.
[28] V. Mandic-Maravic et al., "Sex differences in autism spectrum disorders: does sex moderate the pathway from clinical symptoms to adaptive behavior?", *Scientific Reports*, 5, 2015.
[29] L. Hull, K.V. Petrides, C. Allison, P. Smith, S. Baron-Cohen, M.C. Lai, and W. Mandy, "'Putting on my best normal': social camouflaging in adults with autism spectrum conditions, *Journal of Autism and Developmental Disorders*, 47(8), pp. 2519-2534, 2017.
[30] M.C. Lai et al., "Quantifying and exploring camouflaging in men and women with autism", *Autism*, 21(6), 690-702, 2017.
[31] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scale", in *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain: International Phonetic Association, pp. 774-778, 2003.
[32] A. Nadig and H. Shaw, "Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners", in *Journal of Autism and Developmental Disorders*, *42*(4), pp. 499-511, 2012.