Global and local prosodic entrainment in a task-oriented interaction in autistic and neurotypical children

Joanna Kruyt
Institute of Informatics,
Slovak Academy of Sciences/
in Bratislava, Slovakia
FIIT, Slovak Technical University
in Bratislava, Slovakia
joanna.kruyt@savba.sk

Katarína Polónyiová and **Daniela Ostatníková**
Academic Research Center for Autism,
Institute of Physiology, Faculty of Medicine,
Comenius University in Bratislava, Slovakia

Štefan Beňuš
Institute of Informatics, Slovak Academy of Sciences in Bratislava, Slovakia/
Constantine the Philosopher University in Nitra, Slovakia

Abstract

Entrainment is an inherently interactive phenomenon whereby speakers tend to become more similar over time, for example in their way of speaking or prosody. Prosodic entrainment is associated with positive social measures. Autism spectrum disorder (ASD), on the other hand, is associated with social difficulties, as well as unusual prosody. Taken together, this could suggest that individuals with and without ASD show differences in prosodic entrainment. Such a between-group difference would be a promising direction for automatic, speech-based diagnostic screening tools. The present study aimed to characterise prosodic entrainment in Slovak children with and without ASD using several measurements, to understand participants' entrainment behaviour on multiple levels. Results paint a complex picture of prosodic entrainment in this corpus, but no significant between-group differences in prosodic entrainment were found using both local and global methods, suggesting that prosodic entrainment alone might not be the most promising direction for diagnostic screening tools.

1 Introduction

Early diagnosis of autism spectrum disorder (ASD) is crucial, as it allows for early intervention and can significantly improve quality of life for a child with ASD and their family (e.g. Elder et al., 2017; Fernell et al., 2013). However, getting a formal diagnosis is complicated: waiting lists for assessments are often long, and the process of testing a child and interviewing their family members is time-consuming and requires certified experts. While it is clear that automatic tests cannot replace a comprehensive ASD assessment needed to evaluate and support an individual on the spectrum, any tools that can be used to screen children or assist the diagnostic process could speed up the process and may be especially beneficial to groups that are underdiagnosed with ASD, such as girls (e.g. Gould and Ashton-Smith, 2011; Dean et al., 2017) or children of colour (e.g. Durkin et al., 2017).

Such screening tools can rely on different types of data, but a common way to investigate differences between verbal individuals with ASD and neurotypical peers is centered around speech and language. It must be noted that many tools to distinguish children with ASD from NT children based on language alone usually require the children to have reached a specific level of language development, which not all children with ASD reach. While prosodic entrainment may be determined on the basis of non-verbal vocalisations, this is not commonly done. The present paper will focus on verbal children with ASD.

Existing research has shown that individuals with ASD can be distinguished from individuals without ASD based on their speech: for example, Parish-Morris et al. (2016) correctly identified 68% of children with ASD and 100% of neurotypical (NT) children from a group of children with ASD, children with other (developmental) disorders, and NT children based on their acoustic features and word choices. Acoustic-prosodic features such as speech rate and fundamental frequency (f0) facilitated the distinction between the three groups. Similarly, other studies have used combinations of acoustic features and various machine learning algorithms to distinguish between children with ASD, their NT peers, and children with different developmental disorders (e.g. Lee et al., 2013; Deng et al., 2017-07). In other words, it appears that children with ASD can be distinguished from children

Proceedings of the First ICARD Workshop, pages 1–11

without ASD or those with other developmental disorders based on their speech and language.

A specific phenomenon that is inherently interactive, only occurs during conversation, and can be observed in language is *entrainment* (other terms are also used, such as *alignment*, *convergence*, and *synchrony*) and describes the phenomenon of increased similarity between two interacting people. Entrainment can occur in multiple modalities, such as facial expressions, body movement, and language, and this paper will focus on the latter. Some evidence suggests that individuals with ASD may exhibit or use entrainment differently than NT individuals, specifically at the prosodic level. For example, according to previous research, the prosody of individuals with ASD is often perceived as unusual or "odd", and their pitch can be perceived as unusually flat or overly varied (e.g. Paul et al., 2009; Nadig and Shaw, 2012). Even in the gold standard of ASD diagnostics, the Autism Diagnostic Observation Schedule (ADOS), speech and prosody play a role: for example, inappropriately fast or slow speech, "odd" intonation, unusual pitch and stress, and flat or monotone speech are included as characteristics of ASD (Lord et al., 2008).

These observations may suggest different linguistic entrainment patterns in individuals with ASD, compared to NT individuals. Moreover, linguistic entrainment, specifically at the prosodic level, has been associated with positive social measures such as feelings of closeness and interpersonal rapport (e.g. Michalsky et al., 2018; Pardo et al., 2012; Chartrand and Bargh, 1999). ASD, on the other hand, is associated with marked social difficulties: individuals with ASD report struggling with friendships and romantic relationships more than their NT peers (e.g. Bossaert et al., 2015; Taheri et al., 2016). A possible between-group difference in prosodic entrainment behaviour could in part explain these seemingly unrelated traits.

Further evidence for possible differences in linguistic entrainment between NT and ASD populations comes from research that suggests that individuals with ASD exhibit lower degrees of synchrony or entrainment in other modalities, such as body swaying rhythm (Marsh et al., 2013), pendulum swinging (Fitzpatrick et al., 2018), and finger tapping (Koehne et al., 2016). Reduced entrainment can also be observed in conversational settings: for example, compared to their NT peers, individuals with ASD exhibit less entrainment to their conversational partners in terms of head and hand movements (Noel et al., 2017) and eye blinking patterns (Nakano et al., 2011).

The present paper aims to investigate whether prosodic entrainment differs between verbal children with ASD and their neurotypical peers. Since entrainment is an inherently interactive phenomenon that is associated with positive social measures, a between-group difference in this behaviour is promising for the development of automatic, speech-based diagnostic screening tools. First, prior work on entrainment in these groups will be reviewed.

2 Related work

When it comes to prosodic entrainment in individuals with ASD, a handful of studies have been conducted (see Kruyt and Beňuš (2021) for a review). For example, Ochi et al. (2019) used the "time aligned moving average" method (TAMA by Kousidis et al. (2008), which involves extracting features from a moving window and cross-correlating the resulting time series, to assess entrainment during the diagnostic interview of an ADOS module (Lord et al., 2008). Ochi et al. (2019) reported a range of findings that, taken together, suggest that there may be reduced entrainment in adults with ASD (with a mean age of 26.9 years) compared to their NT peers (mean age of 29.6 years), though it is important to note that all participants in their study were male. While it is common for males to be overrepresented in ASD research (Watkins et al., 2014), the absence of female participants in this study begs the question of whether its results are generalisable to the wider population with ASD.

Lehnert-LeHouillier et al. (2020) elicited task-oriented conversation from participants with (mean age of 12.2 years) and without ASD (mean age of 12.1 years) during a collaborative task and used a geometric approach to quantify entrainment on mean fundamental frequency (f0) and f0 range features by comparing the difference between speakers in the first and last third of every interaction. Results showed that entrainment was observed in conversations with NT participants, but disentrainment was measured in conversations with participants with ASD. There was a significant difference between groups on mean f0 contribution: "on average, participants with ASD contributed 71% less to the convergence in mean f0 during the conversation compared to their NT peers" (Lehnert-

LeHouillier et al., 2020), p. 8). Importantly, results also suggested experimenter's entrainment behaviour differed based on participants: there was a significant interaction effect between age and diagnosis (i.e. ASD or NT) on the experimenter's contribution to the overall observed entrainment, suggesting that age and diagnosis influenced how the experimenter entrained to the participants and that some of the significant between-group differences in entrainment actually originated from the experimenters, rather than the participants.

A comprehensive investigation into entrainment in individuals with (mean age of 19.5 years) and without ASD (mean age of 15.5 years) and their first-degree relatives was conducted by Patel et al. (2022), who found between-group differences in prosodic entrainment: individuals with ASD showed a stronger tendency to dis-entrain rather than entrain, while the opposite was observed in their NT peers. Interestingly, this pattern was mirrored in the parental groups: parents of individuals with ASD (mean age of 50.1 years) also showed prosodic dis-entrainment compared to parents of NT participants (mean age of 47.0 years).

In general, research into prosodic entrainment in individuals with and without ASD is scarce, but results suggest that there may be a pattern of reduced prosodic entrainment in individuals with ASD. However,in each study, different operationalisations and methods are used to measure entrainment: some studies investigated entrainment in smaller time-frames, while others measured it over whole conversations. Similarly, not all studies took into account who was entraining to whom.

In the present study, we will use several previously methods to measure prosodic entrainment, in order to investigate it in different time-frames (i.e. both locally and globally) within the same corpus. Additionally, the age group of our participants differs from the age of participants in existing studies: participants in prior studies were either adults, or slightly older children, and the participants in our study are younger. Comparing the results from our study to those of prior research may shed some light on the development of entrainment behaviours over time.

3 Methods

3.1 Participants and procedure

We collaborated with the Academic Research Center for Autism (ARCA) in Bratislava, Slovakia, and recruited two groups of participants: children with (suspected) ASD (diagnosis was later confirmed with standardised diagnostic tests) and neurotypical (NT) children without suspected ASD or other developmental disorders. All children were native speakers of Slovak and had normal to corrected sight and hearing.

In total, we recruited 67 children (14F, 62M), 41 of which were diagnosed with ASD (7F, 34F) and 35 of which were neurotypical (7F, 28M). The mean age of all recruited children was 9.21 (±1.86) years. For more details on the demographics and test scores of the recruited participants, see Table 1. Note that not all data was included in every analysis due to technical issues.

Table 1: Summary of demographic information and test scores for both groups of participants.

	ASD (7F, 34M		NT (7F, 28M)		t-test
	mean (sd)	range	mean (sd)	range	p
age	9.10 (1.71)	6.14 - 12.30	9.34 (2.04)	6.18 - 12.97	>0.05
IQ	96.80 (16.81)	52 - 131	105.8 (14.84)	67 - 134	<0.05
BRIEF	67.32 (9.61)	47 - 85	58.77 (12.81)	36 - 83	<0.01
ToM	8.29 (2.14)	2 - 14	12.17 (1.74)	8 - 15	<0.001

All children with suspected ASD underwent the full diagnostic procedure, consisting of the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2008) and Autism Diagnostic Interview (Revised, ADI-R). Additionally, the Woodcock-Johnson test was administered to all children to obtain a measure of their intelligence quotient (IQ), and the Comic strip task (Cornish et al., 2010) was used to obtain a theory of mind (ToM) score for each child, since the role of ToM in entrainment is considered a topic of debate. Information regarding the executive functions (EF) of each child was collected using the Behavior Rating Inventory of Executive Function (BRIEF) questionnaire (Gioia et al., 2000) that was filled out by the children's parent(s)/caregiver(s), since executive functioning may also play a role in entrainment. The Slovak versions of all materials were used and all tasks were administered and scored by a certified psychologist.

3.2 Maps task

To elicit conversation, we asked participants to complete the Maps task. In this collaborative task, two participants are given maps that differ slightly. One participant's map has a pre-drawn route, that this participant has to try and describe to their fellow player. This "instruction follower" has to try

and replicate the "instruction giver"'s route as precisely as possible on their own map. The Maps task is commonly used to elicit task-oriented dialogue.

The experimenter who participated in the Maps task with each child was the same clinician who administered the other tests, to ensure that the child was familiar and comfortable with them. The experimenter was thus aware of whether the child received an ASD diagnosis or not. To eliminate the influence of any experimenter bias, she received detailed instructions prior to the experiment. She was aware of the purpose of the study, but primarily knew about a lexical entrainment analysis and knew little about the prosodic analyses.

Typically during a Maps task, roles switch between trials. However, such constant rule-changing can be challenging for individuals with executive functioning difficulties such as many children with ASD. For this reason, we only included one role switch in the Maps task: during the first 3 trials (1 practice trial of which data was not analysed, 2 "real" trials), the child followed instructions, while during the last 3 trials (1 practice, 2 "real"), the child gave instructions.

3.3 Audio recordings and pre-processing

To minimise distress in the children with ASD who often have sensory issues, a directional tabletop microphone and Lavalier microphone were used rather than head-mounted microphones. For reasons of consistency, the experimenter was also recorded with a directional tabletop microphone. Audio recordings were annotated by experience annotators who are native Slovak speakers. Recordings were orthographically transcribed using Transcriber before being transformed to Praat TextGrids that were used to manually convert Transcriber's utterances into IPUs (inter-pausal units) with a silence threshold of 250ms by two trained annotators who were two native speakers of Slovak.

3.4 Analyses

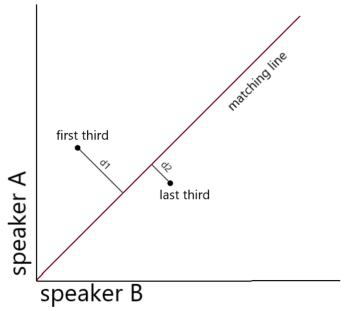
We excluded three files because of technical and audio quality issues, so in total, we used 73 audio files. For each speaker, we used the manually aligned TextGrids to extract median f0 per IPU from the corresponding recordings. We implemented a two-pass filter to avoid pitch tracking errors (Looze and Rauzy, 2009) and used speaker-dependent pitch floors and ceilings: 100–450 Hz for children (Pedersen et al., 2015), and 75-500 for the adult, female experimenter.

3.4.1 Global prosodic entrainment

We used a method introduced by Lehnert-LeHouillier et al. (2020) to measure global prosodic entrainment (see Figure 1), as this allows for the investigation of directionality of entrainment, i.e. how much each speaker entrained to the other speaker. Based on the findings by Lehnert-LeHouillier et al. (2020), who found that the experimenter entrained differently to participants with and without ASD, this was deemed the most appropriate method to analyse our data.

Following Lehnert-LeHouillier et al. (2020), we drew two points that each represented a speaker's median f0 at the start ("first third" in Figure 1) and end of the conversation ("last third" in Figure 1). Next, a "matching line" was drawn to represent the scenario in which both speakers' features are identical. The difference in distance from the start point to this matching line (d1) and the distance from the end point to the matching line (d2) is an indication of the overall entrainment. To calculate each speaker's individual contribution to the overall entrainment, the proportion of change along both the x-axis (speaker B) and the y-axis (speaker A) can be calculated.

Figure 1: Visual representation of the geometric method by Lehnert-LeHouillier et al. (2020), explained in more detail in Section 3.4.1. This figure is inspired by the figure in Lehnert-LeHouillier et al. (2020).



Rather than extracting features from the first and last third of an interaction, as was done by Lehnert-LeHouillier et al. (2020), we calculated median f0 (extracted per IPU and then averaged weighted by IPU duration) for the first two trials and the last two trials of our Maps task. It is important to note that between these trials, the task roles of the participants switched (i.e. the child acted as the instruction follower during the first two trials, and as the instruction giver during the last two trials). The trial/role variable thus represents both a difference in roles as well as in elapsed time.

We investigated whether the overall observed entrainment was predicted by various variables using the following linear mixed effects model formula:

entrainment ∼ group * trial/role + ToM + gender + age + BRIEF + IQ + (1 | participant)

A group * trial/role interaction effect was included since we hypothesised that task role (i.e. instruction giver vs follower) may influence entrainment, and roles in our Maps task switched halfway through. We included the interaction effect to account for the possibility that children from different group react differently to the different roles.

Because prior research by Lehnert-LeHouillier et al. (2020) suggests that differences in observed entrainment between groups may in fact be due to differences in how the adult experimenter interacts with the children, we constructed an additional model using the following formula:

entrainment contribution experimenter ∼ group * trial/role + ToM + gender + age + (1 | participant).

BRIEF and IQ score were not included in this model as we assumed that these factors are not salient or explicit enough to warrant any changes in behaviour of an interlocutor, but the experimenter was aware of the child's diagnosis and their ToM score (as this related to her own research goals as a separate part of this project), so ToM score was kept in the formula. All the LMEMs were implemented using the R package lme4 (Bates et al., 2014) and significance of effects was assessed using lmerTest (Kuznetsova et al., 2017).

3.4.2 Local prosodic entrainment

Additionally, we investigated prosodic entrainment at the local level: we measured proximity, convergence, and synchrony following Levitan et al. (2011-06). This was done to ensure that prosodic entrainment was investigated comprehensively.

Local proximity was determined by using a paired t-test to compare the difference between a target IPU and its adjacent IPU uttered by the other speaker, to the mean difference between the target IPU and 10 random non-adjacent IPUs by the other speaker. Local convergence is calculated conducting a Pearson correlation on the difference between a target IPU and its adjacent IPU uttered by the other speaker, and IPU number as an indication of elapsed time. Local synchrony is measured by performing a Pearson correlation on a target IPU and its adjacent IPU uttered by the other speaker. For local convergence and local synchrony, Pearson correlations are repeated 10x on randomly shuffled data and significant results are only considered valid if less than 2 of these randomly shuffled correlations return significant results.

Note that in the analysis conducted in this experiment, we used speaker-based z-score norming rather than gender-based z-score norming as was used by Levitan et al. (2011-06).

4 Results

4.1 Global prosodic entrainment

The intercept of the model used to predict overall observed prosodic entrainment, at which values correspond to group = NT, trial/role = last two trials, theory of mind score = 0, gender = M, age = 0, IQ = 0, and BRIEF score = 0, is estimated to be 0.91 (see Table 2).

Table 2: Effects in the LMEM constructed for the overall observed prosodic entrainment analysis. No effects were found to be significant.

effect	beta	std	t	df	p
intercept	0.91	9.17	0.10	136	0.922
group	-2.87	2.68	-1.07	136	0.287
trial/role	-2.14	2.11	-1.01	136	0.314
group*trial/role	1.97	3.11	0.63	136	0.528
ToM	0.03	0.42	0.07	136	0.943
gender	-1.59	2.01	-0.79	136	0.430
age	0.18	0.43	0.42	136	0.677
BRIEF	0.02	0.01	0.34	136	0.732
IQ	-0.01	0.05	-0.28	136	0.782

Within this model, all effects were found to be non-significant. Non-significant, negative effects were group, trial/role, gender, and IQ. Non-significant, positive effects were ToM and BRIEF score, age, and the interaction between trial and group (see Table 2).

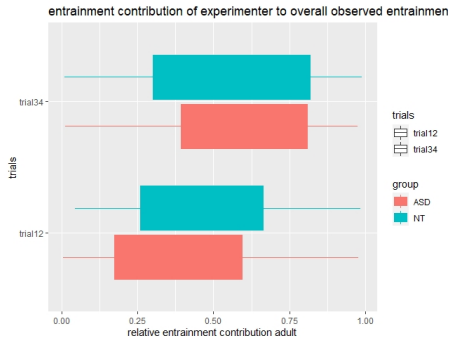
To investigate whether the adult entrained differently to children from either group or during different trial/roles, an additional LMEM was constructed. In this model, the intercept, which corresponds to values of group = NT, trial/role = last two trials, ToM and age, are 0, and gender = M, was estimated to be 0.23 (see Table 3).

In this model, most effects were positive and non-significant, such as group, ToM score, and age. Effects that were negative and non-significant were gender and the interaction between group and trial/role (see Table 3). A significant positive effect

Table 3: Effects in the LMEM constructed for the experimenter's contribution to the overall observed prosodic entrainment analysis. Significant effects are indicated with an asterisk.

effect	beta	std	t	df	p
intercept	0.23	0.16	1.44	138	0.151
group	0.05	0.08	0.55	138	0.581
trial/role	0.18	0.06	2.74	138	0.007*
group*trial/role	-0.14	0.09	-1.45	138	0.151
ToM	0.01	0.01	0.47	138	0.638
gender	-0.05	0.06	-0.87	138	0.386
age	0.02	0.01	1.46	138	0.147

Figure 2: Plot of the experimenter's relative contribution to the overall observed entrainment during different trial/role (trial12 represents the first two real, non-practice trials while trial34 represents the last two trials) and to children from different groups.



of trial/role was found, suggesting that in the second two trials, the experimenter contributed significantly more to the overall entrainment than during the first two trials, and that this happened irrespective of the group membership of the child. This is shown in Figure 2, which suggests that that the experimenter entrained similarly to children with and without ASD regardless of the trial.

4.2 Local prosodic entrainment

An additional set of analyses was conducted to measure entrainment on a different time-scale, namely the local one. Local proximity, convergence, and synchrony were measured, and only a limited number of conversations returned a significant result.

In terms of local proximity, the number of significant measurements of entrainment (N=0) and dis-entrainment (N=1) were equal for the NT and ASD group (see Table 4). A similar finding was obtained for local synchrony, where significant entrainment was not found, and significant dis-entrainment was found in 4 conversations in the NT group, and in 5 conversations in the ASD group (see Table 5).

Table 4: Summary of results for the local proximity measurements. Numbers in cells reflect the number of conversations that returned significant findings.

	sig. entr.	sig. dis.	total sig.	conv.
ASD	0	1	1	39
NT	0	1	1	34
total	0	2	2	73

Table 5: Summary of results for the local synchrony measurements. Numbers in cells reflect the number of conversations (conv.) that returned significant entrainment (sig. entr.) or dis-entrainment (sig. dis.).

	sig. entr.	sig. dis.	total sig.	conv.
ASD	0	5	5	39
NT	0	4	4	34
total	0	9	9	73

The measurement of local entrainment that showed the most substantial between-group differences was local convergence. Using this measurement, significant entrainment was found in 2 conversations in the NT group, and 0 in the ASD group. Note that this is the only local prosodic entrainment measurement that produced significant entrainment results. Moreover, significant dis-entrainment was measured in 7 conversations in the ASD group, and in 4 interactions in the NT group (see Table 6). While this could be taken as an indication of reduced prosodic entrainment and increased prosodic dis-entrainment in the ASD group compared to the NT group, we believe that the fact that the difference between groups is so small, and was only observed on 1 out of 4 measurements, does not indicate substantial, systematic differences between groups when it comes to prosodic entrainment.

Table 6: Summary of results for the local convergence measurements. Numbers in cells reflect the number of conversations that returned significant findings.

	sig. entr.	sig. dis	total sig.	conv.
ASD	0	7	7	39
NT	2	4	6	34
total	2	11	13	73

In short, the tables illustrate that between-group differences are absent or minimal, and while we cannot say with certainty that there is no statistically significant difference between groups in terms of local prosodic entrainment, we have sufficient evidence to believe that the two groups do not consistently and systematically differ in this behaviour.

Interestingly, dyads who showed significant (dis-)entrainment according to one of the local mea-

Table 7: Dyad IDs of the dyads that showed significant entrainment (sig. entr.) or significant dis-entrainment (sig. dis.) according to the three different measures of local entrainment.

proximity				convergence				synchrony			
sig. entr.		sig. dis.		sig. entr.		sig. dis.		sig. entr.		sig. dis.	
ASD	NT	ASD	NT	ASD	NT	ASD	NT	ASD	NT	ASD	NT
		A22	K23		K12	A06	K14			A06	K17
					K17	A16	K16			A08	K26
						A17	K23			A09	K28
						A20	K27			A22	K30
						A21				A41	
						A37					
						A38					

sures of entrainment, often did not show entrainment according to a different measure. Table 7 shows that some dyads, such as A06, A22, and K23, exhibit significant dis-entrainment on more than one dimension. However, for the majority of dyads, this is not the case: most dyads show no significant (dis-)entrainment at all, and out of those that do, most dyads show significant (dis-)entrainment on one measure and exhibit no significant (dis-)entrainment on the other dimensions. In other words, there seems to be no clear pattern in the entrainment patterns of individual dyads.

5 Discussion

The two groups in this experiment did not differ significantly in terms of age. The groups did differ significantly on IQ score, as well as ToM score and BRIEF score (see Table 1). Although we tried to match our NT participants to the children with ASD that we recruited on age and approximate IQ, it is difficult to match two groups of participants perfectly. Our NT group ended up having a significantly higher mean IQ than our group with ASD. Additionally, our NT group performed significantly better on the ToM test. This is in line with findings that suggest that children with ASD may have ToM impairments (e.g. Baron-Cohen, 2000; Baron-Cohen et al., 1985; Tager-Flusberg, 2007).

The same applies to the executive functioning assessment as measured by the BRIEF questionnaire: the NT group had a significantly lower score than the group with ASD, which reflects fewer issues with executive functioning in day-to-day life. Again, this is in line with previous research that shows that ASD is associated with executive functioning challenges (Demetriou et al., 2018).

5.1 Global prosodic entrainment

The results of the geometric prosodic entrainment analysis suggest that group, ToM score, gender, age, trial/role, IQ score, and BRIEF score do not significantly predict the degree of prosodic entrainment that was measured between a child and an experimenter during a Maps task.

Most prior research into prosodic entrainment in individuals with and without ASD suggests that it differs significantly between groups, and often is reduced in the group with ASD. However, in many of these studies, different methods are used to measure entrainment, and most of these methods do not measure each individual's contribution to the overall observed entrainment. A study that shows why this is essential was conducted by Lehnert-LeHouillier et al. (2020), who found a difference between groups in overall observed entrainment, but then found that some of this difference could be explained by the experimenter's behaviour rather than that of the child. While our results neither suggest a significant difference in overall entrainment between groups, nor a significant effects of any predictors on the degree of prosodic entrainment, we nonetheless aimed to investigate whether the experimenter's behaviour towards children from the two groups differed, especially since the experimenter was not naive to the diagnosis of the children she interacted with. For this reason, we constructed an additional statistical model and plotted the experimenter's contribution to the overall observed entrainment in Figure 2, and both suggest that the experimenter entrained similarly to children with and without ASD, which is not in line with findings by Lehnert-LeHouillier et al. (2020).

A possible difference between the experimenter in the present study and in that conducted by Lehnert-LeHouillier et al. (2020) could be their awareness of the children's diagnosis. While the experimenter in our study knew whether a child she was interacting with was diagnosed with ASD or not, we provided her with training in hopes of ensuring that she behaved the same towards all children. It is unclear whether the experimenter(s) in the Lehnert-LeHouillier et al. (2020) study were naive to the children's diagnosis or not, and whether they received training prior to the experiment. This could perhaps, at least in part, explain the differences between the findings of both studies.

5.2 Local prosodic entrainment

An additional set of analyses was conducted to measure entrainment on a different time-scale, namely the local one. Local proximity, convergence, and

synchrony were measured, and only a limited number of conversations returned a significant result.

In terms of local proximity, the number of significant measurements of entrainment (N=0) and dis-entrainment (N=1) were equal for the NT and ASD group (see Table 4). A similar finding was obtained for local synchrony, where significant entrainment was not found, and significant dis-entrainment was found in 4 conversations in the NT group, and in 5 conversations in the NT group (see Table 5).

The measurement of local entrainment that showed the most substantial between-group differences was local convergence. Using this measurement, significant entrainment was found in 2 conversations in the NT group, and 0 in the ASD group. Note that this is the only local prosodic entrainment measurement that produced significant entrainment results. Moreover, significant dis-entrainment was measured in 7 conversations in the ASD group, and in 4 interactions in the NT group (see Table 6). While this could be taken as an indication of reduced prosodic entrainment and increased prosodic dis-entrainment in the ASD group compared to the NT group, we believe that the fact that the difference between groups is so small, and was only observed on 1 out of 4 measurements, does not indicate substantial, systematic differences between groups when it comes to prosodic entrainment.

In short, the tables illustrate that between-group differences are either absent or minimal, and while we cannot say with certainty that there is no statistically significant difference between groups in terms of local prosodic entrainment, we have sufficient evidence to believe that the two groups do not consistently differ in this behaviour.

Interestingly, dyads who showed significant (dis-)entrainment according to one of the local measures of entrainment, often did not show entrainment according to a different measure. Table 7 shows that some dyads, such as A06, A22, and K23, exhibit significant dis-entrainment on more than one dimension. However, for the majority of dyads, this is not the case: most dyads show no significant (dis-)entrainment at all, and out of those that do, most dyads show significant (dis-)entrainment on one measure and exhibit no significant (dis-)entrainment on the other dimensions. In other words, there seems to be no clear pattern in the entrainment patterns of individual dyads.

5.3 General discussion

The results of our global entrainment analyses are not in line with those of Lehnert-LeHouillier et al. (2020) and suggest no significant differences between groups in terms of entrainment or entrainment contribution. Similarly, no consistent differences between groups can be found in any of the local entrainment measurements, despite previous studies suggesting a somewhat consistent pattern of reduced prosodic entrainment in individuals with ASD compared to individuals without ASD. It is possible that this pattern of reduced entrainment only emerges at an older age: almost all participants in previous studies were older than our participants by at least a few years.

We did include age as a fixed effect in our models and found that it did not significantly predict entrainment. It is possible that age alone does not predict entrainment, but that the interaction between age and group membership (i.e. ASD or NT) influences entrainment behaviour: plausibly, as a child grows older, the social interactions they participate in become more complex, which could require more cognitive resources especially for children with ASD who may struggle to navigate some social situations. Increased cognitive effort or load has been associated with decreased entrainment (Abel and Babel, 2017). Future research may aim to investigate how the social skills and abilities of children with and without ASD relate to their entrainment behaviours at different ages.

6 Conclusions

The aim of this study was to investigate prosodic entrainment in children with and without ASD on both the global and local level during the same task. The results of this study paint a complex picture of prosodic entrainment that does not seem to differ systematically between groups of children with or without ASD. More research about prosodic entrainment, its underlying psychological mechanisms, and possible social functions may further elucidate its potential as a screening tool, but at present, empirical evidence is insufficient to develop prosodic entrainment-based screening tools. It is possible that entrainment in other domains, such as body movement, offers a more promising target for easily implementable screening tests, though the role of interaction in such forms of entrainment requires further investigation.

Limitations

Some important limitations of this study relate to experimental design choices we made considering the specific needs of children with ASD. For example, there was a switch in role halfway through the task. While we had good reasons for making this decision, it nonetheless means that task role and elapsed time are always conflated in our analyses (in our trial/role variable). Both of these factors have been shown to influence entrainment (e.g. Reichel et al., 2018), and it is possible that the effects of both cancelled each other out and contributed to the lack of significant results we observed.

The fact that our experimenter was also the clinician who administered other tests to the children similarly may have influenced our results: while we chose to ensure that the children were comfortable with the experimenter to minimise their anxiety and discomfort, the fact that the children were already familiar with the adult may have changed their behaviour towards her. If entrainment is related to social affiliation and rapport building, perhaps our participants felt that enough rapport had been built, and entrainment was no longer necessary. In general, the familiarity between participant and experimenter may make the results of our study more difficult to compare to existing studies.

Finally, we conducted our recordings in the same room where other tests were administered, despite the fact that this room is relatively loud and has an echo. Again, this was done to minimise any distress caused by unfamiliar environments in the children, especially those with ASD.

Another limitation of our study is that only entrainment on median f0 was considered, and no other features such as for example f0 range or speech rate were extracted. While additional analyses can and will be conducted on our collected corpus, it is also important to mention that the recording quality does not allow for the reliable extraction of all features. For example, because we used directional table top microphones rather than head-mounted microphones (again, to minimise distress in children with ASD who often have sensory sensitivities), it is difficult to extract reliable intensity values. Nonetheless, we believe that our collected corpus will be a valuable resource for future research.

Ethics Statement

Ethical approval for both the present experiment as well as the larger overarching research project that this study was part of, was obtained from ARCA's ethics board. Informed consent was obtained from parent(s)/caregiver(s) prior to the experiment. Participants and their parent(s)/caregiver(s) were compensated for their time and participation with gift vouchers.

It is important to note that caution must be taken surrounding the assumption of ToM or social impairments in disorders such as ASD: while traditionally, ASD has been associated with ToM impairments and inherent social deficits, recent empirical evidence suggests that this may be an incorrect assumption (e.g. Paynter et al., 2016; Gernsbacher and Yergeau, 2019). Rather than viewing the communication and ToM and social difficulties of individuals with ASD as their inherent deficits, difficulties with communication and "mind-reading" may be the result of "neurotype mismatches" that occur when an individual with ASD interacts with an NT individual. Individuals with ASD may not lack a theory of mind per se, but rather may struggle to understand the NT mind, specifically. Importantly, this works both ways: NT individuals seem to lack a theory of "autistic" mind (Sheppard et al., 2016; Heasman and Gillespie, 2018). This view is referred to as the "double empathy problem" (Milton, 2012) and is often advocated for by individuals with ASD. It is essential to keep this perspective in mind, especially considering the fact that most existing research into conversation coordination strategies of individuals with ASD was conducted on interactions with a neurotype mismatch.

Acknowledgements

We would like to thank all participants and their parent(s)/caregiver(s) for participating in this study. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 859588, from the Slovak Granting Agency grant VEGA 2/0165/21, and the Slovak Research and Development Agency grant numbers APVV-21-0373 and APVV-20-0070.

References

Jennifer Abel and Molly Babel. 2017. Cognitive load reduces perceived linguistic convergence between

dyads. *Language and Speech*, 60(3):479–502.

S. Baron-Cohen. 2000. Theory of mind and autism: A review. *International review of research in mental retardation*, 23:169–184.

S. Baron-Cohen, A.M. Leslie, and U. Frith. 1985. Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46.

D. Bates, M. Mächler, B. Bolker, and S. Walker. 2014. Fitting linear mixed-effects models using lme4. ArXiv preprint arXiv:1406.5823.

G. Bossaert, H. Colpin, S.J. Pijl, and K. Petry. 2015. Quality of reciprocated friendships of students with special educational needs in mainstream seventh grade. *Exceptionality*, 23(1):54–72.

T.L. Chartrand and J.A. Bargh. 1999. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.

K. Cornish, N. Rinehart, K. Gray, and P. Howlin. 2010. *Comic strip task*. Monash University Developmental Neuroscience and Genetic Disorders Laboratory and Monash University Centre for Developmental Psychiatry and Psychology, Melbourne.

Michelle Dean, Robin Harwood, and Connie Kasari. 2017. The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism*, 21(6):678–689.

E.A. Demetriou, A. Lampit, D.S. Quintana, S.L. Naismith, Y.J.C. Song, J.E. Pye, and A.J. Guastella. 2018. Autism spectrum disorders: a meta-analysis of executive function. *Molecular psychiatry*, 23(5):1198–1204.

J. Deng, N. Cummins, M. Schmitt, K. Qian, F. Ringeval, and B. Schuller. 2017-07. Speech-based diagnosis of autism spectrum condition by generative adversarial network representations. In *Proceedings of the 2017 International Conference on Digital Health*, page 53–57.

Maureen S Durkin, Matthew J Maenner, Jon Baio, Deborah Christensen, Julie Daniels, Robert Fitzgerald, Pamela Imm, Li-Ching Lee, Laura A Schieve, Kim Van Naarden Braun, et al. 2017. Autism spectrum disorder among us children (2002–2010): Socioeconomic, racial, and ethnic disparities. *American journal of public health*, 107(11):1818–1826.

J.H. Elder, C.M. Kreider, S.N. Brasher, and M. Ansell. 2017. Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships. *Psychology research and behavior management*.

E. Fernell, M.A. Eriksson, and C. Gillberg. 2013. Early diagnosis of autism and impact on prognosis: a narrative review. *Clinical epidemiology*, 5:33.

P. Fitzpatrick, J.A. Frazier, D. Cochran, T. Mitchell, C. Coleman, and R.C. Schmidt. 2018. Relationship between theory of mind, emotion recognition, and social synchrony in adolescents with and without autism. *Frontiers in psychology*, 9:1337.

M.A. Gernsbacher and M. Yergeau. 2019. Empirical failures of the claim that autistic people lack a theory of mind. *Archives of scientific psychology*, 7(1):102.

G.A. Gioia, P.K. Isquith, S.C. Guy, and L. Kenworthy. 2000. *Behavior rating inventory of executive function: BRIEF*. Psychological Assessment Resources, Odessa, FL.

Judith Gould and Jacqui Ashton-Smith. 2011. Missed diagnosis or misdiagnosis? girls and women on the autism spectrum. *Good Autism Practice (GAP)*, 12(1):34–41.

B. Heasman and A. Gillespie. 2018. Perspective-taking is two-sided: Misunderstandings between people with asperger's syndrome and their family members. *Autism*, 22(6):740–750.

S. Koehne, A. Hatri, J.T. Cacioppo, and I. Dziobek. 2016. Perceived interpersonal synchrony increases empathy: insights from autism spectrum disorder. *Cognition*, 146:8–15.

S. Kousidis, D. Dorran, Y. Wang, B. Vaughan, C. Cullen, D. Campbell, and E. Coyle. 2008. Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. In *Ninth Annual Conference of the International Speech Communication Association*.

Joanna Kruyt and Štefan Beňuš. 2021. Prosodic entrainment in individuals with autism spectrum disorder. *Topics in Linguistics*, 22(2):47–61.

A. Kuznetsova, P.B. Brockhoff, and R.H.B. Christensen. 2017. lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

H.Y. Lee, T.Y. Hu, H. Jing, Y.F. Chang, Y. Tsao, Y.C. Kao, and T.L. Pao. 2013. Ensemble of machine learning and acoustic segment model techniques for speech emotion and autism spectrum disorders recognition. In *INTERSPEECH*, page 215–219.

H. Lehnert-LeHouillier, S. Terrazas, and S. Sandoval. 2020. Prosodic entrainment in conversations of verbal children and teens on the autism spectrum. *Frontiers in Psychology*, 11:2718.

R. Levitan, A. Gravano, and J. Hirschberg. 2011-06. Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, page 113–117.

C. Looze and S. Rauzy. 2009. Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration. In *Interspeech 2009*.

C. Lord, M. Rutter, P.C. DiLavore, S. Risi, K. Gotham, and S.L. Bishop. 2008. *Autism diagnostic observation schedule (ADOS): Manual*. Western Psychological Services, Los Angeles.

K.L. Marsh, R.W. Isenhower, M.J. Richardson, M. Helt, A.D. Verbalis, R.C. Schmidt, and D. Fein. 2013. Autism and social disconnection in interpersonal rocking. *Frontiers in integrative neuroscience*, 7:4.

J. Michalsky, H. Schoormann, and O. Niebuhr. 2018. Conversational quality is affected by and reflected in prosodic entrainment. In *International Conference on Speech Prosody*, page 389–392.

D.E. Milton. 2012. On the ontological status of autism: the 'double empathy problem'. *Disability Society*, 27(6):883–887.

A. Nadig and H. Shaw. 2012. 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*, 42(4):499–511.

T. Nakano, N. Kato, and S. Kitazawa. 2011. Lack of eyeblink entrainments in autism spectrum disorders. *Neuropsychologia*, 49(9):2784–2790.

J.P. Noel, M.A. Niear, N.S. Lazzara, and M.T. Wallace. 2017. Uncoupling between multisensory temporal function and nonverbal turn-taking in autism spectrum disorder. *IEEE Transactions on cognitive and developmental systems*, 10(4):973–982.

K. Ochi, N. Ono, K. Owada, M. Kojima, M. Kuroda, S. Sagayama, and H. Yamasue. 2019. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PloS one*, 14(12):0225377.

J.S. Pardo, R. Gibbons, A. Suppes, and R.M. Krauss. 2012. Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1):190–197.

J. Parish-Morris, M. Liberman, N. Ryant, C. Cieri, L. Bateman, E. Ferguson, and R.T. Schultz. 2016. Exploring autism spectrum disorders using hlt. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, page 74. NIH Public Access.

Shivani P Patel, Jennifer Cole, Joseph CY Lau, Gabrielle Fragnito, and Molly Losh. 2022. Verbal entrainment in autism spectrum disorder and first-degree relatives. *Scientific reports*, 12(1):11496.

R. Paul, S.M. Orlovski, H.C. Marcinko, and F. Volkmar. 2009. Conversational behaviors in youth with high-functioning asd and asperger syndrome. *Journal of autism and developmental disorders*, 39(1):115–125.

J.M. Paynter, D. Keen, and V.J. Rose. 2016. Systematic review documents limited empirical support for the practical application of the theory of mind model of asd. *Evidence-Based Communication Assessment and Intervention*, 10(3-4):131–139.

M Pedersen, ABA Agersted, and A Jønsson. 2015. Aspects of adolescence and voice: Girls versus boys–a review. *J Child Adolesc Behav*, 3(211):2.

U.D. Reichel, Š. Beňuš, and K. Mády. 2018. Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication*, 100:46–57.

E. Sheppard, D. Pillai, G.T.L. Wong, D. Ropar, and P. Mitchell. 2016. How easy is it to read the minds of people with autism spectrum disorder? *Journal of autism and developmental disorders*, 46(4):1247–1254.

H. Tager-Flusberg. 2007. Evaluating the theory-of-mind hypothesis of autism. *Current directions in psychological science*, 16(6):311–315.

A. Taheri, A. Perry, and P. Minnes. 2016. Examining the social participation of children and adolescents with intellectual disabilities and autism spectrum disorder in relation to peers. *Journal of Intellectual Disability Research*, 60(5):435–443.

E.E. Watkins, Z.J. Zimmermann, and A. Poling. 2014. The gender of participants in published research involving people with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 8(2):143–146.