Journal of Autism and Developmental Disorders (2019) 49:2572–2580
https://doi.org/10.1007/s10803-019-03905-5

BRIEF REPORT

Brief Report: Acoustic Evidence for Increased Articulatory Stability in the Speech of Adults with Autism Spectrum Disorder

Mikhail Kissine[1] · Philippine Geelhand[1]

Published online: 1 February 2019

Abstract

Subjective impressions of speech delivery in Autism Spectrum Disorder (ASD) as monotonic or over-precise are widespread but still lack robust acoustic evidence. This study provides a detailed acoustic characterization of the specificities of speech in individuals with ASD using an extensive sample of speech data, from the production of narratives and from spontaneous conversation. Syllable-level analyses (30,843 tokens in total) were performed on audio recordings from two sub-tasks of the Autism Diagnostic Observation Schedule from 20 adults with ASD and 20 pairwise matched neuro-typical adults, providing acoustic measures of fundamental frequency, jitter, shimmer and the first three formants. The results suggest that participants with ASD display a greater articulatory stability in vowel production than neuro-typical participants, both in phonation and articulatory gestures.

Keywords Autism · Prosody · Acoustics · Voice quality · F0 · Formants · Jitter · Shimmer

While socio-pragmatic, communicative difficulties constitute a core component of diagnostic definitions of Autism Spectrum Disorder (ASD), language impairment is currently considered as a specifier of the disorder (American Psychiatric Association 2013, pp. 90–91). Linguistic profiles vary greatly both across the spectrum and individuals with ASD's life-span, with around 60–70% of individuals on the spectrum reaching, at one stage or another, fully functional structural language (e.g. Kim et al. 2014). However, speech abnormalities constitute an extremely robust characteristic of ASD, already present in the very first descriptions of the disorder (Asperger 1944; Kanner 1946). A great proportion of individuals diagnosed with autism show atypical prosodic development (e.g. Peppé et al. 2006); these abnormalities tend to persist into adulthood, and are independent of improvement in other areas of language (DePape et al. 2012; Shriberg et al. 2001; Fusaroli et al. 2017). Even adults with linguistic and cognitive levels fully within typical ranges are frequently described as having unexpectedly flat and monotonous prosody or over-precise diction (e.g. Attwood 2015, pp. 230–231). These phonetic characteristics clearly contribute to the social communication difficulties faced by individuals with autism (e.g. DePape et al. 2012), and are likely to negatively impact the quality of their social interactions, hindering the development of social-communicative abilities (Bone et al. 2015; Warlaumont et al. 2014). Furthermore, even subtle acoustic characteristics, such as atypical nasal resonance, may correlated with lower socio-communicative skills (Paul et al. 2005).

Earlier studies have relied on perceptual methods to characterize abnormalities in speech production, resulting in subjective descriptions of acoustic patterns in ASD as flat, monotone, variable, sing-songy, pedantic, machine-like, stilted, bizarre or exaggerated (Baltaxe and Simmons 1985; Lord et al. 1994). An obvious drawback of perceptual judgments is that they are probably not sufficiently reliable to be considered as clinically meaningful. For example, gold-standard diagnostic instruments such as the Autism Diagnostic Observation Schedule, Second edition (ADOS-2; Lord et al. 2012), include speech abnormalities in their criteria and are coded during administration; yet, they are not considered in the diagnostic algorithm, as there is not enough subjective agreement between clinicians (Bone et al. 2015).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10803-019-03905-5) contains supplementary material, which is available to authorized users.

✉ Mikhail Kissine
mkissine@ulb.ac.be

[1] ACTE at LaDisco & ULB Neuroscience Institute, Université libre de Bruxelles, CP 175, avenue F.D. Roosevelt, 1050 Brussels, Belgium

In order to improve the diagnostic utility and discriminatory power of the patterns of speech production in autism, research is increasingly focusing on more objective measures and analyses of acoustic features of speech delivery. Mean fundamental frequency (F0) and F0 variability emerge as yielding the most reliable differences between individuals with ASD and comparison participants (Fusaroli et al. 2017). Somehow puzzlingly, though, most studies report a higher F0 variation in participants with ASD than in neuro-typical (NT) comparison groups (Green and Tobin 2009; Diehl et al. 2009; Bonneh et al. 2011; Grossman et al. 2013; Filipe et al. 2014). That is, while there is little reason to doubt the validity of widespread reports of monotonic speech delivery in autism, they do not reflect in the currently available acoustic evidence. The main objective of this paper is to explore in more depth the acoustic—and hence articulatory—features that may be responsible for this impression, and thus pave the way to reconciling acoustic evidence with subjective reports.

Most of the existing studies analyse acoustic features at the level of individual words, utterances or entire narratives (but see Paul et al. 2008; Bone et al. 2015). Yet, it is possible that while people with ASD exhibit increased pitch variation at such macro levels, they also display reduced F0 variation the level of individual syllable nuclei. Zooming on the syllable level also provides an opportunity to assess whether speech delivery is associated with reduced variation in the relative frequency (*jitter*) or amplitude (*shimmer*) of the vibration of vocal folds. While abnormally high jitter and shimmer are usually associated with the presence of speech pathologies, it is also possible that speech in ASD is characterised by an overly regular phonation, which may create an impression of monotonic delivery. Furthermore, analysing vowels allows to gather the values of the first three formants (F1–F3), which are strongly correlated with the shape of supra-glottic articulators. Therefore, the extent of variation of F1–F3 across different realisations of the same vowel by the same speaker provides a quantitative measure into this speaker's articulatory stability.

In this paper, we test the hypothesis that speech delivery in adults with ASD is characterised by an increased invariance, both at the level of phonation and at the level of articulators' position. Using an extensive sample of data, both from adults with ASD and pairwise matched neuro-typical (NT) adults, we provide a very detailed acoustic characterisation of the specificities of autistic speech. Our hope is that this description will contribute to a better understanding of the often reported impressions of abnormal tone of voice in individuals with ASD. We recorded two tasks of the entire ADOS session by 20 adults with ASD and pairwise-matched NT adults, from which we extracted and coded all V, CV, VC and CVC syllables (30843 in the total final sample).[1] In addition to syllable duration, we analysed the median F0 (Hz), the F0 range (in semi-tones), as well as jitter, shimmer and the values of F1, F2 and F3. These latter measures were used to compute an *F1–F3 dispersion* index, understood as the Euclidean distance on the three-dimensional F1, F2 and F3 space.

Focusing on fine articulatory characteristics is also crucial to gather new insights into the factors that underlie atypical speech delivery in autism. Atypical prosody is often linked to difficulties individuals with ASD experience in adapting their conversational contribution to the context, and in conveying meaning through supra-segmental components. However, verbal individuals with ASD are capable of relying on prosody to grasp a large array of linguistic information, such as ambiguous phrase boundaries, focus structure or the contrast between assertions and questions (Peppé et al. 2007; Chevallier et al. 2011). Likewise, Paul et al. (2005) found no difference in the perception and production of pragmatic and affective prosody between young adults with ASD and NT participants. It is therefore possible that the peculiarities of speech in autism are not fully accounted for by deficient access to the linguistic functions of prosody, but also owe to a distinctive, overly precise, execution of articulatory gestures. Of particular relevance for this issue is the nature of the tasks employed to elicit speech production. Some studies use picture naming (Bonneh et al. 2011; Nakai et al. 2014) or reading tasks (Green and Tobin 2009). Arguably, such data are fairly constrained and are probably not entirely representative of speech delivery in every-day situation. Other authors use narrative retelling to obtain more natural samples (Diehl et al. 2009; Grossman et al. 2013; Bone et al. 2015). But while more natural, narrative retelling is still different from a genuinely spontaneous verbal production. Furthermore, building a coherent narrative is specifically challenging for individuals with autism (e.g. Baixauli et al. 2016; Stirling et al. 2014), which, again, may impact on speech quality during the task. To avoid these potential biaises, in this study, we analysed data both from a narrative task and a more natural exchange on the topic of solitude. In this way, we should be able to control whether some acoustic features in participants with ASD are induced by the nature of the speech elicitation context.

[1] V = vowel, C = consonant.

Table 1 Experiment 1: descriptive statistics for participants per task

Task	Group	N	Mean age (sd)	*t*	FSIQ (sd)	*t*	VIQ (sd)	*t*	ADOS	*t*
Narrative	ASD	20	28.1 (11.48)	0.04	110.95 (27.14)	− 0.07	112.45 (21.33)	1.03	10.45 (3.12)	12.45***
	NT	20	27.9 (11.53)		111.45 (15.19)		111.00 (11.93)		1.0 (1.34)	
Solitude	ASD	18	28.9 (11.7)	0.03	113.22 (24.88)	0.13	116.11 (16.06)	1.03	10.39 (3.43)	10.91***
	NT	18	28.8 (11.74)		112.39 (12.58)		111.39 (10.96)		0.94 (1.3)	

FSIQ full-scale IQ, *VIQ* Verbal IQ, as measured by WAIS-IV (Wechsler 2016), *ADOS* total ADOS scores

$p < 0.001$***

Methods

Participants

Participants in the present study came from a pool of French-speaking participants recruited within a larger project on discourse in autism. Participants in the ASD group were recruited via ACTE register of volunteers and from a specialised secondary school for adolescents with ASD. The clinical group was selected as having autism conforming to the criteria of the DSM-IV. Neuro-typical (NT) participants in the comparison group were recruited via advertisements on the internet. Inclusion criteria for both groups included: (1) age between 15 and 60 years, (2) a Full-Scale IQ (FSIQ) score above 70, (3) Verbal IQ (VIQ) score above 70 and (4) normal or corrected-to-normal vision and audition. For the participants in the comparison group, a further inclusion criterion was the absence of known psychiatric, developmental or neurological disorder. Participants were matched pairwise on age (± 1 year difference) and gender. A total of 24 participants in each group were recruited.

The present study is based on the audio recordings of two sub-tasks from the ADOS-2 (Lord et al. 2012): *Telling a story from a picture book*, henceforth *Narrative* and *Solitude*. Data were not available, due to technical or experimental issues, for four participants with ASD in the Narrative task and six participants for the Solitude task. When data were not available for a participant with ASD, we also excluded from the analyses data for the corresponding, pairwise matched NT participant. The final sample included 20 participants (13 male, 7 female) per group—mean FSIQ: 110.95 (ASD) and 111.45 (NT)—in the Narrative task, and 18 (13 male, 5 female) participants per group for the Solitude task—mean FSIQ: 113.22 (ASD) and 112.39 (NT). Table 1 provides descriptive statistics for the participants per task.

Material and Data Collection

The material used in this study comes from the administration of ADOS-2 (Lord et al. 2012) by an accredited ADOS assessor. ADOS-2 is a valid source of conversational data as Modules 3 and 4 are built almost entirely on conversation, and the cues of the different tasks approximate natural conversational situations. Furthermore, the semi-structured quality of these modules created comparable situations across participants. The entire ADOS-2 was video-taped, and converted into an audio file. For ease of analysis, the entire audio file was subdivided into the various tasks of the ADOS-2. The study reported here is based on the Narrative and Solitude tasks.

Both tasks were administered during the standard ADOS-2 procedure in a quiet room. For the Narrative task, the 29-page picture book *Tuesday* (Wiesner 1991) was used to elicit a narrative from each participant. The book is considered wordless with only four sentences providing temporal indications (‘Tuesday evening, around eight’ at the beginning of the story; ‘11:21 pm’ and ‘4:38 am’ during the story; and ‘Next Tuesday, 7:58 pm’ on the last page). *Tuesday* is about frogs that are suddenly able to fly on their lily pads, and start exploring the neighbourhood, surprising those still awake. The experimenter interviewing the participants introduced the task by saying: ‘This is a wordless picture book, I will start telling you the story, and you will finish it’. The two first story boards were told by the experimenter and the rest was told by the participant. When necessary, the experimenter provided some encouragement to pursue the story with prompts in the form of backchannelling (‘mhm’) or general comments (‘Tell me more’, ‘And here, what’s happening?’). For the Solitude task the experimenter followed the ADOS-2 guidelines and asked about the participant’s understanding of and opinion on the concept of solitude.

Data Preparation

The audio files for each tasks were then processed in Praat (Boersma and Weenink 2018). The audio recordings were first orthographically transcribed in Praat by research assistants trained in linguistic data transcription. Based on this orthographic transcription, a phonetic transcription was created using the phonetization function of the Praat plug-in EasyAlign (Goldman 2011). The second author manually checked the output of the phonetization function. These two transcriptions were then segmented into words and syllables

Table 2 Experiment 1: analyzed syllables by voice quality measure, per group and task

Group	Task	Median F0	F0 range	Jitter	Shimmer
ASD	Narrative	7881 (0.75)	7980 (0.75)	7564 (0.71)	7147 (0.68)
	Solitude	3713 (0.81)	3766 (0.82)	3502 (0.77)	3243 (0.71)
NT	Narrative	9165 (0.72)	9319 (0.78)	8723 (0.69)	8019 (0.63)
	Solitude	2072 (0.77)	2108 (0.78)	1926 (0.71)	1706 (0.63)
Total		22831 (0.75)	23173 (0.76)	21715 (0.71)	20109 (0.66)

The proportion this number represents relative to the total of syllables kept for the acoustic analysis is given between brackets

using the Phone segmentation function of EasyAlign. These two segmentations were manually checked again to make sure the boundaries of each word and syllable were correctly aligned to the speech in the audio. The result of this procedure was a TextGrid file for each participant's audio recording with four tiers, namely (1) an orthographic transcription; (2) a phonetic transcription; (3) a word tier and (4) a tier corresponding to syllable segmentation. For the purpose of the present study, this latter syllable tier was then manually checked (again by the second author) to remove the speech of the experimenter and only keep the syllables produced by the participant. Syllables corresponding to inaudible speech were also removed: 54 (participants with ASD) and 28 (NT participants) syllables in the Narrative task and 10 (participants with ASD) and 3 (NT participants) syllables in the Solitude task. Finally, we also excluded syllables contained in overlapping speech between the participant and the experimenter: 287 (participants with ASD) and 76 (NT participants) syllables in the Narrative task and 13 (participants with ASD) and 49 (NT participants) syllables in the Solitude task.

Acoustic Analyses

All acoustic analyses were performed using Praat (Boersma and Weenink 2018). French has no diphtongs; however, some syllables onsets may consist of the combination between a consonant and a liquid [l], a rhotic ([ʀ] or [ʁ]) or a glide ([j], [w] or [ɥ]). The acoustic properties of such complex syllable onsets, as well as those of CC onsets, may complicate the analysis. For this reason, only V, CV and CVC syllables were kept for the syllable analysis. A total of 30,843 syllables were eventually submitted to acoustic analysis.

For some syllables, reliable measures of voice quality could not always be generated, resulting in partial loss of data. However, as can be seen from Table 2, the remaining sample is still very large. One-way ANOVAs did not show any group difference on the proportion of data loss for median F0 ($F(1,42) = 2.42;p = 0.13$), F0 range ($F(1,42) = 1.72;p = 0.19$), jitter ($F(1,42) = 2.38;p = 0.13$) and shimmer ($F(1,42) = 3.03, p = 0.09$); likewise, there was no task difference in data loss for median F0 ($F(1,2) = 4.68;p = 0.16$), F0 range ($F(1,2) = 5.56;p = 0.14$), jitter ($F(1,2) = 2.15;p = 0.28$) and shimmer ($F(1,2) = 13.56;p = 0.06$).

In order to minimise potential effects of co-articulation and target the most stable period of the vowel nucleus, all measures of voice quality were computed only at the interval set between the 0.25 and the 0.75 of the duration of each syllable. For each participant, we first computed the maximum and the minimum F0 using the auto-correlation method (Boersma 1993) with the time-step set at 0.75/*pitch-floor*; the range was set at 75–800 Hz for female participants and at 75–400 Hz for male ones. If the minimum or the maximum F0 obtained were equal to the range limits of the auto-correlation method, we computed again these values decreasing the minimum values 5 Hz or increasing the maximum values by 5 Hz until the the maximum and the minimum F0 were strictly comprised within the range thus set. Next, using the maximum and the minimum values thus obtained and the same method, we computed, for each speaker and each syllable the median F0 and the F0 range in semi-tones. In order to compare median F0 and F0 range on syllable nuclei and on entire words, these two measures were also collected per word.

We also computed local jitter and local shimmer, again at the 0.25–0.75 of each syllable duration. The minimum F0 values and the standard Praat settings were used in each case. Jitter refers to the variation in the frequency of phonation cycle from cycle, while shimmer refers to the variation in amplitude of phonation cycle from cycle. For each participant, we then collected the maximum value of the fifth formant, using Burg analysis (window length of 25 ms and pre-emphasis from 50 Hz), with five formants and the default maximal value of 5500 Hz for female participants and 5000 Hz for male participants. If the maximum value thus obtained was equal to the maximum set, we reran the analysis increasing the maximum value by steps of 25 Hz until the maximum formant frequency obtained was strictly inferior to the maximum thus set. Next, we computed, for each speaker and each syllable the median values for the first three formants (F1, F2 and F3). For each type of vowel V and participant, we computed the mean value of the median values of the first three formants: V_{mF1}, V_{mF2} and V_{mF3}. Then, for each participant and each syllable S with the vowel V and participant, we computed the Euclidean distance between the median values of the first three formants of S, S_{F1}, S_{F2} and S_{F3} and the mean formant values of V for this participant:

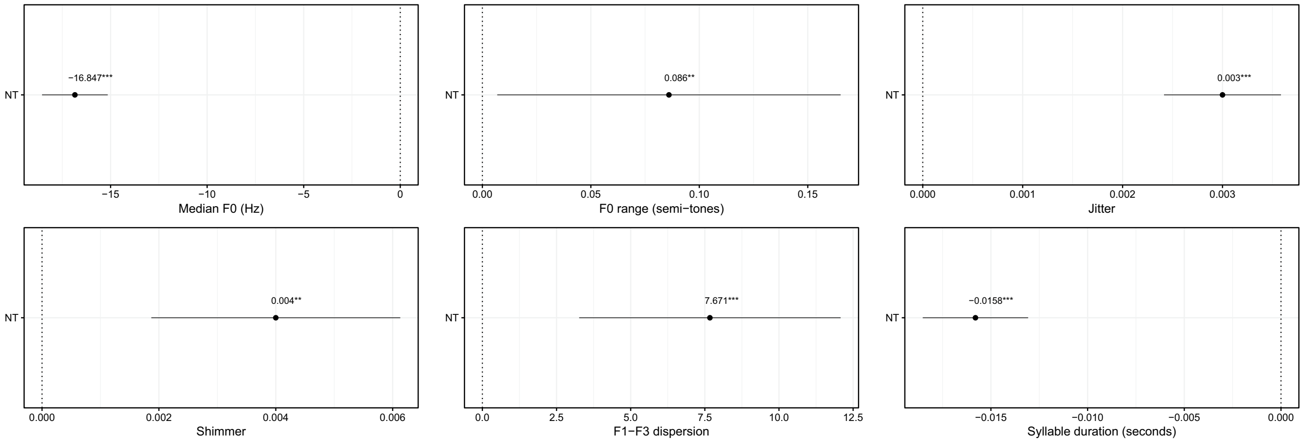

Fig. 1 Caterpillar plots of group effects on acoustic measures. Horizontal bars represent 95% CIs; the ASD group is the intercept; $p < 0.001$***; $p < 0.01$**

Table 3 Robustness checks for the group effect on median F0

	(1)	(2)	(3)	(4)
Group NT	− 16.85*** (0.87)	− 20.15*** (0.73)	− 10.78*** (1.21)	− 9.85*** (1.18)
Gender—male		− 73.61*** (0.76)	− 66.14*** (1.08)	− 65.69*** (1.05)
Group NT × gender male			− 14.75*** (1.51)	− 14.79*** (1.47)
Age				− 1.18*** (0.03)
Number of observations	22,831	22,831	22,831	22,831

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

$\sqrt{(V_{mF1} - S_{F1})^2 + (V_{mF2} - S_{F2})^2 + (V_{mF3} - S_{F3})^2}$. This distance gives the *F1–F3 dispersion index* for each syllable.

Statistical Analyses

All statistical analyses on syllables are performed by implementing multi-level linear regressions in R (R Core Team 2016), using in the lme4 package (Bates et al. 2015). Models used to assess group effects include by-syllable random intercepts; those used to assess task effects include task by-syllable and by-participant random slopes. Group and task variables were dummy-coded. All the group effects, discussed in detail below, are summarised in Fig. 1. Significance of fixed effects was assessed by performing likelihood ratio tests relative to a model with an identical in random effect structure, but without the effect at hand. Post hoc comparisons of least square-means, with Tukey adjustment for multiple comparisons and Satterthwaite method for estimating degrees of freedom, were implemented in the lsmeans package (Lenth 2016). Analyses at the word level were performed using linear regressions.

Results

As can be seen from Fig. 1, which displays group effects for each acoustic variable, there were significant differences between participants with ASD and NT participants on each of the aspects we investigated. We begin by detailing these results, as well as the corresponding robustness checks, variable by variable. Next, we will turn to task effects.

Hierarchical multilevel regressions revealed a strong group effect on median F0 ($\chi^2(1) = 373.55$; $p < 0.001$), with median F0 being higher in the ASD group ($\beta = 16.85$; $se = 0.86$). Since F0 is usually lower in male voices, we controlled this result for gender. As shown in Table 3, the group effect persisted, even when checks for gender—and age—were added. To further explore the gender effect, we ran log-likelihood model comparisons, which revealed the expected gender effect ($\chi^2(1) = 7877.4$; $p < 0.001$). The group × gender interaction also proved significant ($\chi^2(2) = 94.625$; $p < 0.001$). Post hoc comparisons revealed that all contrasts of this interaction were significant (all $p < 0.001$). Median F0 values were significantly higher in the ASD group, for both male ($\beta = 25.52$; $se = 0.91$; $p < 0.001$) and female ($\beta = 10.77$; $se = 1.2$; $p < 0.001$) participants.

There was a group effect on F0 range (in semi-tones) ($\chi^2(1) = 4.58$; $p = 0.032$), with the F0 range being higher in the NT group ($\beta = 0.086$; $se = 0.004$). There was also a highly significant group effect on jitter ($\chi^2(1) = 109.77$; $p < 0.001$), with higher jitter values in the NT group ($\beta = 0.31 \times 10^{-2}$; $se = 0.029 \times 10^{-2}$). Likewise, there was a group effect on shimmer ($\chi^2(1) = 10.55$; $p = 0.001$), with higher values in the NT group ($\beta = 0.4 \times 10^{-2}$; $se = 0.01 \times 10^{-2}$). It could be that higher F0 range, jitter and shimmer in the NT group are simply a mechanic consequence of lower F0 in this group. For this reason we controlled groups effects on F0 range, jitter and shimmer, by adding median F0 as a fixed factor; we also controlled these results for gender and age. As can be seen from Tables 4, 5 and 6 the group effect subsisted in each case (all $p < 0.001$).

Interestingly, while measures of median F0 collected on words went in the same direction ($\beta = -16.23$; $se = 0.76$; $p < 0.001$), the difference in F0 range failed to reach significance ($\beta = 0.07$; $se = 0.04$; $p = 0.099$), which underscores the advantage of finer-grained level of acoustic analysis.

Hierarchical multilevel regressions also revealed a strong group effect on the F1–F3 dispersion index

Table 4 Robustness checks for the group effect on F0 range

	(1)	(2)	(3)	(4)
Group NT	0.09* (0.04)	0.20*** (0.04)	0.22*** (0.04)	0.22*** (0.04)
Median F0		0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)
Gender—male			0.20*** (0.05)	0.22*** (0.05)
Age				0.01** (0.00)
Number of observations	23,173	22,831	22,831	22,831

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

Table 5 Robustness checks for the group effect on jitter

	(1)	(2)	(3)	(4)
Group NT	$10^{-2} \times 0.31$*** ($10^{-2} \times 0.03$)	$10^{-2} \times 0.42$*** ($10^{-2} \times 0.03$)	$10^{-2} \times 0.52$*** ($10^{-2} \times 0.03$)	$10^{-2} \times 0.52$*** ($10^{-2} \times 0.03$)
Median F0		$10^{-2} \times 0.01$*** ($10^{-4} \times 0.02$)	$10^{-2} \times 0.01$*** ($10^{-4} \times 0.03$)	$10^{-4} \times 0.01$*** ($10^{-4} \times 0.03$)
Gender−male			$10^{-2} \times 0.83$*** ($10^{-2} \times 0.04$)	$10^{-2} \times 0.89$*** ($10^{-2} \times 0.04$)
Age				$10^{-2} \times 0.02$*** ($10^{-4} \times 0.14$)
Number of observations	21,715	21,714	21,714	21,714

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

Table 6 Robustness checks for the group effect on shimmer

	(1)	(2)	(3)	(4)
Group NT	$10^{-2} \times 0.35$** ($10^{-2} \times 0.1$)	$10^{-2} \times 0.39$*** ($10^{-2} \times 0.1$)	$10^{-2} \times 0.54$*** ($10^{-2} \times 0.1$)	$10^{-2} \times 0.53$*** ($10^{-2} \times 0.1$)
Median F0		$10^{-4} \times 0.3$** ($10^{-4} \times 0.1$)	$10^{-4} \times 0.7$*** ($10^{-4} \times 0.1$)	$10^{-4} \times 0.81$*** ($10^{-4} \times 0.1$)
Gender—male			0.01*** ($10^{-2} \times 0.1$)	0.01*** ($10^{-2} \times 0.1$)
Age				$10^{-3} \times 0.26$*** ($10^{-4} \times 0.5$)
Number of observations	20,109	20,109	20,109	20,109

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

Table 7 Robustness checks for the group effect on F1–F3 dispersion

	(1)	(2)	(3)
Group NT	7.67*** (2.25)	7.66*** (2.25)	8.22*** (2.24)
Gender—male		− 8.34*** (2.34)	− 7.89*** (2.34)
Age			1.15*** (0.10)
Number of observations	30,483	30,483	30,483

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

Table 8 Robustness checks for the group effect on syllable duration

	(1)	(2)	(3)
GroupNT	− 0.02*** ($10^{-2} \times 0.14$)	− 0.02*** ($10^{-2} \times 0.14$)	− 0.02*** ($10^{-2} \times 0.14$)
Gender—male		$10^{-2} \times -0.18$ ($10^{-2} \times 0.14$)	$10^{-2} \times -0.17$ ($10^{-2} \times 0.14$)
Age			$10^{-2} \times 0.01$ ($10^{-2} \times 0.01$)
Number of observations	30,483	30,483	30,483

$p < 0.001$***; $p < 0.01$**; $p < 0.05$*

($\chi^2(1) = 11.63$; $p < 0.001$). Consistently with voice quality measures, F1–F3 dispersion was greater in the NT group ($\beta = 7.67$; $se = 2.25$). Finally, there was a strong group effect on syllable duration ($\chi^2(1) = 128.2$; $p < 0.001$), syllable duration being significantly shorter in the NT group ($\beta = 0.16 \times 10^{-2}$; $se = 0.14 \times 10^{-2}$). As can be seen in Tables 7 and 8, both effects subsist when controlled for gender and age.

There was no effect of task on median F0 ($p = 0.78$) or on F0 range ($p = 0.33$). There was a task effect on jitter ($\chi^2(2) = 11.01$; $p = 0.001$)—with higher jitter in the Solitude task ($\beta = 0.3 * 10^{-2}$; $se = 0.08 * 10^{-2}$)—, but no task × group interaction ($p = 0.32$). Likewise, there was a task effect on shimmer ($\chi^2(2) = 7.9$; $p = 0.005$)—with higher values in the Solitude task ($\beta = 0.9 * 10^{-2}$; $se = 0.32 * 10^{-2}$)—, but no task × group interaction ($p = 0.4$). There was also a task effect on F1–F3 dispersion ($\chi^2(2) = 5.57$; $p = 0.018$), with higher dispersion in the Solitude task ($\beta = 14.73$; $se = 6.1$). Again, there was no task × group interaction ($p = 0.5$). Finally, there was a task effect on syllable duration ($\chi^2(1) = 27.74$; $p < 0.001$), as well as a task × group interaction ($\chi^2(2) = 7.91$; $p = 0.019$). Overall syllable duration was longer in the Narrative task ($\beta = 0.27 * 10^{-2}$; $se = 0.53 * 10^{-2}$; $p < 0.001$). In the Narrative task, syllable duration was longer for the ASD group ($\beta = 0.16 \times 10^{-2}$; $se = 0.59 \times 10^{-2}$; $p = 0.042$), but not in the Solitude task ($p = 0.25$).

Discussion

The results of our extensive acoustic analysis at the syllable level yield a rather consistent picture of the acoustic specificity of speech delivery in adults with ASD. Starting with voice quality, as can be seen from Fig. 1, NT participants displayed greater variation on the F0 range, phonation frequency (jitter) and amplitude (shimmer). All these spectral measures thus indicate a greater stability of voicing during vowel production in adults with ASD. Furthermore, even though median F0 values were also significantly higher in the ASD group—and consistently so across tasks—, lower F0 and spectral variability in the ASD group cannot be explained away by higher average pitch. These results strongly suggest, therefore, that adults with ASD display less variability in the vibration of vocal folds during vowel production. A remarkably similar conclusion can be drawn from our data on formant dispersion. Recall that this index refers, for each participant, to the Euclidean distance between the median values of the first three formants of a given vowel and the corresponding average values for this vowel for this participant. This dispersion index emerged as significantly higher in the NT group. Since the values of the first three formants are strongly determined by the position of articulators during vowel production, this result shows that participants with ASD produce vowels with more invariant articulatory gestures than NT adults.

Interestingly, while the acoustic measures converge towards a greater articulatory stability in participants with ASD, syllable duration was also longer in this group, at least in the Narrative task. We will return to task effects in a moment, but it is worth emphasising that, a priori, a longer articulation of the vowel trivially leaves more room for variation. That both spectral and formantic measures go in the opposite direction in participants with ASD, independently of longer syllable duration, is a further indication of the robustness of increased articulatory stability in their speech.

On the one hand, these results are in line with the widespread subjective reports of monotonic or over-precise speech delivery in ASD. While there certainly must be other factors—such as the interaction between prosody and information structure—driving such an impression, our acoustic data contributes to a more objective understanding of it. On the other hand, however, they are also in contradiction with some of the previous literature, in particular with studies that, contrary to us, found increased pitch variation in their participants with ASD (Green and Tobin 2009; Bonneh et al. 2011; Grossman et al. 2013; Filipe et al. 2014; Diehl et al. 2009).

One potential reason for this discrepancy is probably related to the task used to collect the acoustic data. For instance, the acoustic analyses in Diehl and Paul (2012) and Filipe et al. (2014) rely on the expressive part of the PEPS-C (Peppé and McCann 2003), in which the participants have to produce words (16 experimental items) with the intonation corresponding to the scene depicted in a vignette (e.g. a child is offering some food vs is looking at it in a bowl). Likewise, Bonneh et al. (2011) had children with ASD name pictures for a period of 60 s (with an average of 27 words for the ASD group). To begin with, the total of acoustic tokens thus obtained is fairly limited. Furthermore, producing single words in such contexts is quite removed from naturally occurring speech, and children or adults with ASD may find this task challenging; all these factors may contribute to higher stress and more variable pitch. Other studies that report greater F0 variation in participants with ASD include a reading component, which may induce a greater articulatory variability than what occurs in natural speech. For instance, Green and Tobin (2009) combined a reading task with elicited semi-spontaneous speech; Grossman et al. (2013) used a task in which participants had to retell a short video-taped story, while relying on a written vignette that contained the exact wording of the story to be retold (the authors aimed at investigating the capacity to reproduce the correct intonation). Other authors used a narrative retelling, based on a video-tape, to elicit spontaneous speech (Diehl et al. 2009). However, as already mentioned in the introduction, building a coherent narrative is also known to present difficulties for people with ASD (Baixauli et al. 2016; Stirling et al. 2014; Banney et al. 2015; Canfield et al. 2016)—in fact, this is the reason why it features in the ADOS-2 (Lord et al. 2012).

In that relation, it is interesting to note that our results indicate that the Narrative task, in general, induces a greater stability in spectral and formantic measures than the Solitude task, suggesting the more spontaneous nature of the latter. Recall also that the only interaction between the group and task factors we found was due to the fact that syllable duration was longer for the ASD group in the Narrative task, again suggesting that it is particularly challenging for participants with ASD. An advantage of our study is to use data both from online narrative production and a more spontaneous exchange on the topic of solitude, thus allowing a more diverse and extensive set of data.

Another obvious difference between our study and the literature just evoked relates to the age group. While the studies cited above investigated children or young adolescents, we analysed data from adults. It is of course possible that speech in younger individuals with ASD is at first characterised by a greater pitch range, to stabilise in the opposite direction during adulthood. More likely, however, the reason why, contrary to previous studies, we found greater acoustic invariance in our participants with ASD owes to the nature of our acoustic analyses. Acoustic studies on ASD usually limit themselves to the word level or to predefined equal slices of the speech signal. While these methods are certainly less time-consuming than the analysis syllable by syllable we conducted, working on a finer-grained level allowed us to zoom on acoustic data that have clear articulatory correlates.

Of course, the increased articulatory stability this study uncovered exhausts neither the prosody in autism nor the way speech by individuals with ASD is perceived by neurotypicals. Our participants with ASD represent only a subgroup of individuals on the spectrum. Clinical sub-groups may differ on subtle acoustic measures, so that our results do not necessarily generalise across the entire autism spectrum. Future research should seek to replicate these results with different languages and contexts, and try to link them both with prosodic contours and perception studies.

An intriguing question we have to leave open for the moment is that of the causes of this articulatory stability. One explanation would be a difference in register, such that participants with ASD favour an overprecise diction in contexts where neuro-typicals speak in a more relaxed way. If so, one should find interactional contexts in which no group difference, of the kind we reported here, would arise. It is worth noting, however, that in our data the Solitude task gave rise to less stable acoustic measures in both groups than the Narrative task. The absence of group × task interaction may be seen as an indication that while articulatory invariance in ASD is influenced by the nature of the linguistic activity, it also persists as a peculiarity of the speech delivery in ASD independently of the context. Another possibility could be that individuals with ASD pay more care to the exact performance of target articulatory gestures, whereas some of this precision is lost for neuro-typicals—for reasons, again, to be uncovered in future studies.

Acknowledgments We thank the acting editor and three anonymous reviewers for very useful comments on previous versions of this paper. The present research has been made possible thanks to the F.R.S.-FNRS Research Incentive Grant F.4502.15 to Mikhail Kissine and a Foundation Jean-François Peterbroeck doctoral grant to Philippine Geelhand. This generous support is acknowledged. We thank all our participants and their families for their time and willingness to take part in our study.

Compliance with Ethical Standards

Conflict of interest All authors declare no conflict of interest.

Ethical Approval All procedures in this study were approved by the ethics committee of Erasme Hospital in accordance with the 1964 declaration of Helsinki and its later amendments.

Informed Consent All adult participants provided informed consent. Adolescent participants provided informed assent with their parents providing informed consent.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Arlington: American Psychiatric Association.

Asperger, H. (1944). Die "autistischen psychopathen" im kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, *117*(1), 76–136.

Attwood, A. (2015). *The complete guide to Asperger's syndrom*. London: Jessica Kingsley.

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, *59*, 234–254.

Baltaxe, C. A. M., & Simmons, J. Q. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. Mesibov (Eds.), *Communication problems in autism* (pp. 95–125). Boston, MA: Springer.

Banney, R. M., Harper-Hill, K., & Arnott, W. L. (2015). The autism diagnostic observation schedule and narrative assessment: Evidence for specific narrative impairments in autism spectrum disorders. *International Journal of Speech-Language Pathology*, *17*(2), 159–171.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, *17*, 97–110.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. Amsterdam: Institute of Phonetic Sciences.

Bone, D., Black, M. P., Ramakrishna, A., Grossman, R., & Narayanan, S. S. (2015). Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism. In *Interspeech* 2015 (pp. 1616–1620).

Bonneh, Y., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*, 237.

Canfield, A. R., Eigsti, I.-M., de Marchena, A., & Fein, D. (2016). Story goodness in adolescents with autism spectrum disorder (ASD) and in optimal outcomes from ASD. *Journal of Speech, Language, and Hearing Research*, *59*(3), 533–545.

Chevallier, C., Noveck, I., Happé, F. G. E., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, *49*(3), 507–517.

DePape, A. M. R., Chen, A., Hall, G. B. C., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology*, *3*, 1–13.

DePape, A. M. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory processing in high-functioning adolescents with autism spectrum disorder. *PLoS ONE*, *7*(9), e44084.

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(1), 123–134.

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, *44*(8), 1972–1981.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407.

Goldman, J.-P. (2011). Easyalign: An automatic phonetic alignment tool under Praat. In *Interspeech 2011*.

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult... but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315.

Grossman, R. B., Edelson, L. R., & Tager-Flusberg, H. (2013). Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *56*(3), 1035–1044.

Kanner, L. (1946). Irrelevant and metaphorical language in early infantile autism. *American Journal of Psychiatry*, *103*, 242–246.

Kim, S. H., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Language and communication in autism. In F. R. Volkmar, S. J. Rogers, R. Paul, & K. A. Pelphrey (Eds.), *Handbook of autism and pervasive developmental disorders* (4th ed., pp. 230–262). Hoboken: Wiley.

Lenth, R. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule-2nd edition (ADOS-2)*. Los Angeles, CA: Western Psychological Corporation.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain and Development*, *36*(6), 516–522.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 861.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4–5), 345–354.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*(10), 1776–1791.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115.

Stirling, L., Douglas, S., Leekam, S. R., & Carey, L. (2014). The use of narrative in studying communication in autism spectrum disorders: A review of methodologies and finding. In J. Arciuli & J. Brock (Eds.), *Communication in autism* (pp. 171–217). Amsterdam: John Benjamins.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324.

Wechsler, D. (2016). *Echelle d'intelligence de Wechsler pour enfants et adolescents - WISC V*. Paris: ECPA.

Wiesner, D. (1991). *Tuesday*. Boston: Clarion Books.