RESEARCH ARTICLE

Phonetic Inflexibility in Autistic Adults

Mikhail Kissine, Philippine Geelhand, Marie Philippart De Foy, Bernard Harmegnies, and Gaétane Deliens

This study examined whether the atypical speech style that is frequently reported in autistic adults is underpinned by an inflexible production of phonetic targets. In a first task, 20 male autistic adults and 20 neuro-typicals had to read and produce native vowels. To assess the extent to which phonetic inflexibility is due to an overall fine-grained control of phonetic behavior or to a lack of flexibility in the realization of one's phonological repertoire, the second task asked participants to reproduce artificial vowel-like sounds. Results confirmed the presence of a greater articulatory stability in the production of native vowels in autistic adults. When instructed to imitate artificial vowel-like sounds, the autistic group did not better approximate the targets' acoustic properties relative to neuro-typicals but their performance at reproducing artificial vowels was less variable and influenced to a greater extent by the articulatory properties of their own vocalic space. These findings suggest that the greater articulatory stability observed in autistic adults arises from a lack of flexibility in the production of their own native vowels. The two phonetic tasks are devoid of any pragmatic constraint, which indicates that phonetic inflexibility in autism is partly independent of register selection. ***Autism Res** 2021, 00: 1–11*. © 2021 International Society for Autism Research, Wiley Periodicals LLC.

Lay Summary: Autistic and neuro-typical adults took part in two tasks: one in which they produced vowels from French, their native tongue, and the other where they imitated unfamiliar vowels. Autistic adults displayed significantly less variation in their production of different French vowels. In imitating unfamiliar vowels, they were more influenced by the way they pronounce French vowels. These results suggest that the atypical speech style, frequently attested in autistic individuals, could stem from an unusually stable pronunciation of speech sounds.

Keywords: autism; language; phonetic compliance; acoustics; prosody; phonetic inflexibility in autistic adults

Introduction

Atypical and delayed language development constitutes one of the most frequent features of Autism Spectrum Disorder (ASD). However, around 60% of individuals on the spectrum eventually reach functional language, and a significant proportion of them enter adulthood with linguistic skills within a typical range [Kim, Paul, Tager-Flusberg, & Lord, 2014]. Yet, even these verbal adults use language in a perceptibly atypical way. Alongside difficulties in adopting the speaker's perspective or constructing a coherent narrative [Deliens, Papastamou, Ruytenbeek, Geelhand, & Kissine, 2018; Geelhand, Deliens, Papastamou, & Kissine, 2020; Geurts, Kissine, & van Tiel, 2020], an unusual speech delivery style is frequently reported in verbal autistic adults. Speech in autistic individuals is often perceived as monotone, machine-like, stilted, bizarre or exaggerated [Baltaxe & Simmons, 1985; Lord, Rutter, & Le Couteur, 1994]. Better understanding atypicalities in the speech of autistic individuals is of paramount importance, as impressions of atypical prosody can hinder opportunities to develop social relationships during everyday conversations, affecting peers' perceptions of the speaker, as well as their overall experience of the social interaction [Boyd et al., 2016; Paul et al., 2005; Shriberg & Widder, 1990].

Most of the studies that attempted to objectivize these subjective descriptions focused on acoustic characteristics of prosody. Yet, this body of research failed to uncover any robust acoustic characteristics that would account for a monotonic or mechanic speech style, as most studies reported a *higher* variability in autistic speakers, especially at the level of the fundamental frequency (F0) [see Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008; Filipe, Frota, Castro, & Vicente, 2014; Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017; Green & Tobin, 2009; Grossman, Edelson, & Tager-Flusberg, 2013; Wehrle, Cangemi, Hanekamp, Vogeley, & Grice, 2020]. Furthermore,

From the ACTE, Université libre de Bruxelles, Bruxelles, Belgium (M.K., P.G., G.D.); ULB Neuroscience Institute, Université libre de Bruxelles, Bruxelles, Belgium (M.K., P.G., G.D.); Center for Research in Linguistics (LaDisco), Université libre de Bruxelles, Bruxelles, Belgium (M.K., P.G.); Laboratoire de Phonétique, Institut de Recherche en Sciences et Technologies du Langage, Université de Mons, Mons, Belgium (M.P.D.F., B.H.); Center for Research in Cognition and Neurosciences (CRCN), Université libre de Bruxelles, Bruxelles, Belgium (G.D.)
Received August 28, 2020; accepted for publication January 7, 2021
Address for correspondence and reprints: Mikhail Kissine, CP 175, Université libre de Bruxelles, 50 av. F.D. Roosevelt, 1050 Bruxelles, Belgium. E-mail: mkissine@ulb.ac.be
Published online 00 Month 2021 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.2477

while some subtle aspects of the linguistic functions of prosody may be impaired in ASD, very few significant group differences have been reported, be it in comprehension or in production [Chevallier, Noveck, Happé, & Wilson, 2011; Paul et al., 2005; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007; Scheerer, Shafai, Stevenson, & Iarocci, 2020].

Lack of robust supra-segmental correlates of the atypical speech delivery in ASD may owe to the nature and the variety of tasks used to elicit verbal production, some of which autistic participants may find challenging for independent reasons [see Kissine & Geelhand, 2019, for a detailed discussion]: picture naming [Bonneh, Levanon, Dean-Pardo, Lossos, & Adini, 2011; Nakai, Takashima, Takiguchi, & Takada, 2014], reading [Green & Tobin, 2009] or narrative retelling [Bone, Black, Ramakrishna, Grossman, & Narayanan, 2015; Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Filipe et al., 2014; Grossman et al., 2013]. However, it is also possible that more reliable group differences are located at segmental articulatory levels. Kissine and Geelhand [2019] analyzed a large corpus of naturally produced speech and, based on the analysis of multiple acoustic characteristics, found that there was less variation in the production of different vowel types by autistic speakers relative to their neuro-typical (NT) peers. This result suggests that subjective impressions of a monotonic, flat tone of voice in many verbal autistic speakers could partly be due to increased articulatory stability—or, equivalently, to a lack of phonetic flexibility—in the production of articulatory gestures. In addition to potentially providing an objective, acoustic explanation to subjective impressions triggered by autistic speech, the presence of such atypical articulatory stability may lead to new insights into the apprehension and acquisition of language in autism.

Life-long socio-communicative deficits constitute a core characteristic of ASD, and it would be therefore natural to ask whether overly inflexible articulatory production in autism could be due to difficulties in adapting one's delivery style to a more informal register. Data analyzed by Kissine and Geelhand [2019] came from two ADOS [Lord et al., 2012] tasks: production of a narrative based on an illustrated book and semi-structured conversation on the topic of solitude. The former task is more formal and gave rise to less variation in vowel productions than the latter—in both autistic and NT participants. This result suggests an influence of register, that is, a systematic linguistic variation as a function of situational features. However, while more formal registers are associated with greater articulatory stability, autistic participants also exhibited significantly more stable vowels in both narrative production and semi-structured conversation, that is, independently of the task register. Nonetheless, it remains possible that Kissine and Geelhand's [2019] autistic participants were overall adopting a more formal register, so that lower phonetic flexibility was due to a difficulty in switching to a more informal register rather than being a feature inherent in their speech production.

The first objective of this paper is to confirm the presence of phonetic inflexibility in verbal adults with ASD using a more controlled paradigm, in which potential register confounds are neutralized. Our expectation is that even in a task that requires mere production of one's native language vowels—thus devoid of any pragmatic constraints—autistic participants will display less articulatory variation.

If, as we expect, autistic participants do display lower phonetic flexibility in such a task, this feature of autistic speech has to be explained in terms at least partly independent of register selection. One possibility, then, would be that lower phonetic flexibility in ASD owes to an overall hyper-precise mapping of phonetic targets on sensory-motor commands, that is, independently of the nature of the phonetic target. Such a hypothesis would be in line with models that associate ASD with an enhanced processing of low-level, local properties of perceptual stimuli [Happé & Frith, 2006; Mottron, Dawson, Soulières, Hubert, & Burack, 2006; Pellicano & Burr, 2012], such as, for instance, superior pitch discrimination [Bonnel et al., 2010]. Another, not necessarily incompatible explanation is that the atypically high articulatory stability documented by Kissine and Geelhand [2019] in autistic individuals is limited to the sounds of their native tongue, and can thus be described as an atypically inflexible realization of the phonological repertoire. According to the first line of explanation, autistic speakers would display an enhanced control of a wide range of articulatory gestures; according to the latter they should be heavily constrained in the way they produce their native vowels. Delvaux, Huet, Piccaluga, and Harmegnies [2014] introduced the concept of "phonetic compliance," defined as the aptitude to reproduce with high fidelity a wide variety of speech sounds independently of their (dis)similarity with one's native phonological repertoire. High phonetic compliance thus corresponds to an increased ability to reach a phonetic target, while minimizing the influence of one's native language.

The experimental paradigm Delvaux et al. [2014] devised to test this concept (used with NT participants) is therefore particularly suited for adjudicating between the two explanations of phonetic inflexibility in ASD that we have just evoked.The part of the test by Delvaux et al. [2014] that we borrow here combines a controlled production of native vowels with a task that requires participants to reproduce artificial vowel-like sounds. These artificial sounds are generated within a formantic space, within value ranges found in natural language vocalic spaces, but not necessarily overlapping with the participant's native language vocalic space. If the articulatory stability in the production of native vowels by autistic adults is due to an atypically high phonetic compliance, i.e. to a hyper-precise realization of any type of articulatory target, one should expect such hyper-precision to also surface with non-native targets, so that productions by autistic participants should better approximate artificial vowel-like sounds than those by NT participants. If, by contrast, articulatory inflexibility in ASD is limited to the individual's exemplars of native vocalic space, autistic participants should not perform better than NT participants in the imitation of non-native sounds. A further advantage of Delvaux et al.' [2014] paradigm is that it allows us to assess the attraction participants' own vocalic spaces exert on their production of non-native sounds. If articulatory stability in ASD is due to overprecise articulatory gestures, their production of non-native sounds should not be influenced by the articulatory properties of their own vocalic space. By contrast, if articulatory stability in ASD is driven by a less flexible association between articulatory gestures and native phonological targets, it seems reasonable to expect their production of artificial vowels to be more influenced by the acoustic properties of their production of native vowels.

The main acoustic correlates of vowel production are given by the first three formants (F1, F2, and F3). These formants correspond to the first three resonant frequencies of the vocal tract, and are thus determined by the position of supra-glottic articulators. Roughly, F1 is negatively correlated with vowel height, while F2 is correlated with vowel anteriority; F3 can be partly predicted from F1 and F2, but is also correlated with lip rounding [Johnson, 2012]. Variation on the F1–F3 space thus represents a good approximation of articulatory variation. That is, lower within-participant formant variation across different tokens of native vowel types in autistic participants relative to NT participants would be an indication of a greater articulatory stability in the former group. While Kissine and Geelhand [2019] used the first three formants in their analyses of articulatory stability in vowel production, Delvaux et al. [2014] excluded F3 from most of their metrics. We used F1, F2, and F3 in this paper; however, the significance of all the effects reported below remains unaffected when only the first two formants are kept in the analyses. The fundamental frequency (F0), resulting from the frequency of vocal fold vibration, is also included within acoustic analyses reported below. Kissine and Geelhand [2019] reported that autistic participants had an overall higher F0 than NT participants [also Filipe et al., 2014]. However, in Kissine and Geelhand [2019], vowel productions by autistic participants also displayed a lower variation in F0 range relative to NT participants. This latter result somehow contradicts other studies which found higher F0 variation in ASD, albeit at the level of words or entire utterances [Bonneh et al., 2011; Diehl et al., 2009; Filipe et al., 2014; Grossman et al., 2013]. Reduced F0 range may constitute another indication of an increased articulatory stability in ASD. It is therefore important to determine whether Kissine and Geelhand's [2019] result generalizes to a more controlled task.

Summing up, we expect to find lower variation in F1–F2–F3 values and in the F0 range of the native vowels produced by autistic *vs* NT participants. This result would confirm higher articulatory stability in ASD, reported by Kissine and Geelhand [2019]. If such higher articulatory stability is due to an overall higher articulatory precision, the

F1–F2–F3 values of attempts at reproducing artificial vowels should be closer to the F1–F2–F3 values of the targets in autistic participants. If, by contrast, higher articulatory stability in ASD owes to a more constrained production of the native phonological repertoire, on the one hand, no such group difference should emerge, and, on the other hand, the formant values of the attempts at reproducing artificial sounds should reflect stronger attraction of the participant's native vowel space in autistic participants. Finally, we also expect F0 to be overall higher in autistic participants.

Methods

Participants

The 40 male, native speakers of French who participated in this study, 20 autistic adults and 20 NT adults, were a subset of participants from a larger project. All autistic participants (mean age ± SD = 31.55 ± 10.70; range 17–52) were recruited from the ACTE register of volunteers and held a previous diagnosis of ASD from a multidisciplinary team, based on the Autism Diagnostic Observation Schedule [ADOS; Lord et al., 2012]. The NT group (29.05 ± 8.98 years old; range 17–52) was recruited *via* advertisements on the internet. All participants were native French speakers and had normal or corrected-to-normal vision and audition; additionally, no NT participant had a history of developmental delays, psychiatric diagnoses or neuro-cognitive impairments.

All participants had average to above average IQ score on the Wechsler Adult Intelligence Scale [WAIS-IV; Wechsler, 2008] except one autistic participant (IQ = 52). Statistical analyses with and without this participant's data revealed no differences, therefore the data were included in all analyses. There was no difference between groups in educational (Mann-Whitney $U = 181$; $P = 0.61$) and economical background ($t(37) = -1.74$; $P = .09$), as assessed by an adapted version of the family affluence scale [Hobza, Hamrik, Bucksch, & De Clercq, 2017; Torsheim et al., 2016], in age ($t(38) = 0.80$; $P = 0.43$) or Total IQ ($t(38) = -0.88$; $P = 0.38$). As expected, the

ASD group had higher scores on the Autism spectrum Quotient [AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001] and lower score on the Empathy Quotient [EQ; Baron-Cohen & Wheelwright, 2004] than the NT group ($t(38) = 10.92$; $P < 0.001$ and $t(38) = -4.81$; $P < 0.001$, respectively). Demographic and psychometric data are presented in Table 1. All participants agreed to participate in the current study after reading and signing an informed consent form approved by the local Biomedical Ethics Committee (CUB Hôpital Erasme, Université Libre de Bruxelles); approval number P2018/414.

Tasks

The native vowel production task (Task 1) and the artificial vowel imitation task (Task 2) were both implemented *via* the E-Studio component of the E-PRIME 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA). The E-run component was used for the administration of the tasks. In the native vowel production task, 10 French oral vowels (/a/, /œ/, /i/, /u/, /e/, /ɛ/, /ə/, /o/, /ɔ/, /y/) were displayed one by one in their written form, using the most frequent French grapheme corresponding to each vowel (<a>, <eux>, <i>, <ou>, <é>, <è>, <e>, <au>, <o>, <u>) in white font against black background. Participants first read the following instruction: "In this task, some letters will be displayed on the screen. You are asked to produce these sounds out loud as if you were speaking." Each stimulus was presented for 2000 ms with an inter-stimulus interval of 250 ms. Five replications of each of the 10 vowels were made, one in each block of 10 trials. The stimulus order within the 5 blocks was randomized. In the artificial vowel imitation task, four blocks of 94 synthesized vocoids borrowed from Delvaux et al. [2014] were auditorily delivered. All stimuli were built using a Klatt synthesizer [Klatt, 1980], lasted 200 ms, with a constant F0 contour of 110–90 Hz and are equally distributed over a F1–F2–F3 space in mels; see Delvaux et al. [2014] for a detailed presentation of the stimuli. Stimuli were presented during 2000 ms with an inter-stimulus interval of 250 ms and were randomized across 4 blocks. Participants were instructed to reproduce these synthesized sounds as closely as possible, as if these were speech sounds. Subjects were given a short break between each block.

Table 1. Participant Statistics

Measure	ASD (n = 20)	NT (n = 20)
Age (years)		
Mean (SD)	31.55 (10.7)	29.05 (8.98)
Range	17–52	17–52
Total IQ		
Mean (SD)	112.06 (22.8)	117.9 (14.18)
Range	52-151	90-141
Educational status		
Mean (SD)	1.63 (1)	1.18 (1)
Range	1–6	1–5
Economic status[a]		
Mean (SD)	7.21 (2.55)	8.4 (1.64)
Range	3–13	6–11
AQ		
Mean (SD)	35 (7.33)	13.9 (4.59)
Range	20–46	4–21
EQ		
Mean (SD)	22.25 (9.23)	33.70 (5.29)
Range	6–39	26–42
ADOS total score[b]		
Mean (SD)	11.00 (4.95)	
Range	6–20	

[a]Missing data for one ASD participant.
[b]Missing data for four ASD participants.

Procedure

Participants were fitted with a headset mounted with a microphone (Model Trust Gaming GXT 10). This headset system allowed to keep the distance between the subject's mouth and the microphone constant. The native vowel production task (Task 1) was always administered before the artificial vowel imitation task (Task 2).

Data Preparation

All acoustic analyses were performed using Praat [Boersma & Weenink, 2018]. For each participant and each task, we first computed the maximum and the minimum F0 using the auto-correlation method [Boersma, 1993] with the time-step set at 0.75/*pitchfloor*; as all our participants were male, the initial range was set at 75-400 Hz. 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 or increasing the maximum values by 5 Hz steps, until the maximum and the minimum F0 were strictly comprised within the range thus set. Next, using the maximum and the minimum values obtained and the same method, we computed, for each speaker and each syllable the median F0 and the F0 range in semi-tones.

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 set at 5000 Hz. 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 item the median values for the first three formants (F1, F2 and F3). For each type of vowel V and participant, we then computed the mean value of the median values of the first three formants: $VmF1, VmF2$ and $VmF3$.

The acoustic data obtained allowed us to compute *distance from target*, *dispersion* and *attraction* indexes. Figure 1 provides illustrative spectrograms for the production of native vowels in Task 1. For Task 1, a *dispersion* index of the production of native vowels on the F1-F3 space was computed, for each participant and each vowel, as the Euclidean distance between the first two formant value of each token of the vowel and the corresponding mean formant value for all the productions of this vowel by the speaker at hand [Delvaux et al., 2014; Karlsson & van Doorn, 2012; Kissine & Geelhand, 2019]. 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}. For each participant, given the production P of the vowel V, the median values of the first three formants of P, P_{F1}, P_{F2}, and P_{F3}:

$$\text{disperstion (native vowels)} = \sqrt{(P_{mF1} - V_{F1})^2 + (P_{mF2} - V_{F2})^2 + (P_{mF3} - V_{F3})^2}$$

Figure 2 provides illustrative spectrograms for one target of Task 2 (imitation of artificial vowels), along with four attempts to reproduce it by a participant. In Task 2, the *distance from target*, for each participant and each attempt A at imitating the target T, was defined as the Euclidean distance between the median values of the first three formants of A, A_{mF1}, A_{mF2} and A_{mF3}, and the first three formants of T, T_{F1}, T_{F2}, and T_{F3}:

$$\text{distance from target} = \sqrt{(A_{mF1} - T_{F1})^2 + (A_{mF2} - T_{F2})^2 + (A_{mF3} - T_{F3})^2}$$

We also computed an *attraction index*, in order to assess the extent to which participants' production of artificial vowels was influenced by their own vocalic space. Using data from Task 1 (native vowels production), we computed, for each participant and vowel V, the mean values of the first three formants. Next, for participant and each attempt A in Task 2, we computed the Euclidean distances between the first three formants of A, and the mean values of the first three formants for each native vowel for this participant; then we determined the vowel with the lowest distance from A, $min(dist_V)$. Finally, if the *distance from target* for the A was superior to $min(dist_V)$, the attraction index was set at 1 and at 0 otherwise. In informal terms, the attraction index was 1 if the participant's attempt was closer to at least one acoustic value of his native vowel space than to the target.

Finally, we also computed a *dispersion index* in Task 2 in the same way as in Task 1. For each participant, given the attempt A at the production of the target T, we computed the dispersion index as the Euclidean distance between

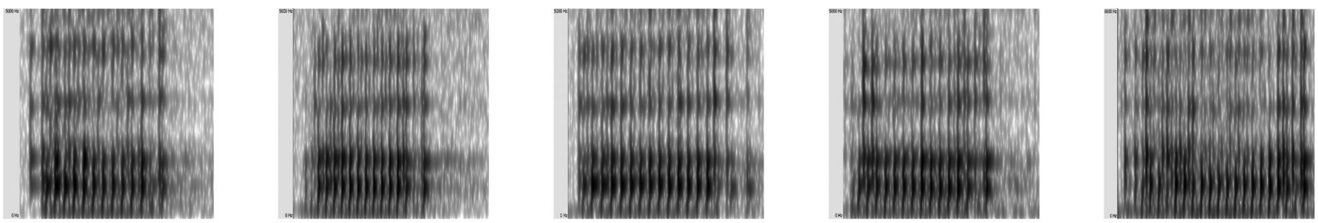

Figure 1. Task 1: Spectrograms of five productions (V_{1-5}) of the French vowel /a/ by a neuro-typical participant. Darker horizontal lines correspond to formants; closer values of the first three formants across productions indicate higher articulatory stability.

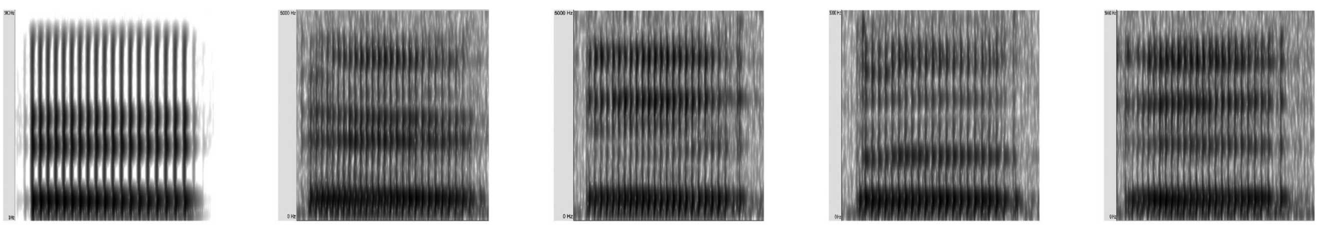

Figure 2. Task 2: Artificial vowel imitation task. From left to right: spectrograms of a target stimulus (T) and of four attempts (A_{1-4}) at reproducing it by a neuro-typical participant. Darker horizontal lines correspond to formants; lower distance between the first three formants of an attempt and those of the target indicates better imitation of the target.

the first three formants of A and the mean values of the first three formants of all the attempts at the production of T for this participant.

Data Analysis

All statistical analyses were performed by implementing multi-level linear or logistic regressions in R [R Core Team, 2016], using in the lme4 package [Bates, Maechler, Bolker, & Walker, 2015]. All models included a random by-item intercept; significance of the Group effects was assessed by performing likelihood ratio tests relative to a model with an identical random effect structure, but without the effect at hand. Because one of the aims of the paper was to confirm the findings by Kissine and Geelhand [2019], whenever relevant, we explicitly compared the effect sizes to those in Kissine and Geelhand [2019].

Results

Task 1: Native Vowel Production

The addition of the fixed Group effect significantly improved the model fit for median F0 ($\chi_2(1) = 98.29$; $P < 0.001$) and for F0 range ($\chi_2(1) = 8.32$; $P = 0.004$). In line with Kissine and Geelhand [2019], see Figure 3A & B, autistic participants displayed higher median F0 values ($\beta = 14.18$, $se = 1.41$), but a narrower F0 range ($\beta = -0.83$; $se = 0.29$). Turning to the *dispersion* index for the production of native vowel, the addition of the fixed Group effect also significantly improved the model fit ($\chi_2(1) = 14.54$; $P < 0.001$). In line with Kissine and Geelhand [2019], the dispersion index was significantly higher in the NT group ($\beta = 44.91$, $se = 11.76$); see Figure 3C.

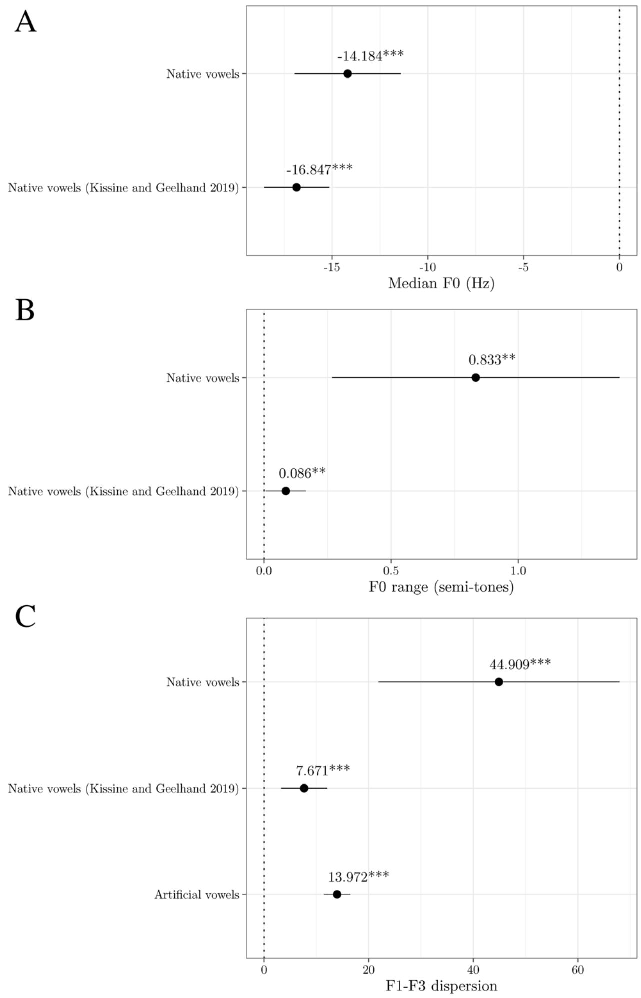

Figure 3. Caterpillar plots of group effects on acoustic measures; (**A**) Median F0; (**B**) F0 range; (**C**) Dispersion on the F1–F3 space. Horizontal bars represent 95% CIs; the ASD group is the intercept; $^{***}P < 0.001$; $^{**}P < 0.01$.

Task 2: Artificial Vowel Imitation

The addition of the fixed Group effect significantly improved the fit for the *distance from target* ($\chi_2(1) = 6.84$; $P < 0.001$), indicating that NT participants' attempts to reproduce the target artificial vowel were significantly closer to the target's first three formants than the attempts by participants with ASD ($\beta = -9.12$; $se = 3.49$). Recall, however, that in Task 1, autistic participants were found to have an overall higher F0 in their production of native vowels; see Figure 3A. Since the F0 of the artificial vowels was relatively low (110–90 Hz), it could be that the task was made more difficult for autistic participants, who, in addition to attempt to match the articulatory position of the target, needed to modulate their F0 to a greater extent than NT participants. In order to control for this confound, we added the median F0 value of the attempt within the model that predicted the *distance from target*. The addition of the median F0 significantly increased the overall model fit ($\chi_2(1) = 258.51$; $P < 0.001$), but annihilated the group effect ($\beta = 3.36$, $se = 3.54$; $P = 0.343$). In sum, autistic and NT participants do not seem to significantly differ in their ability to match the articulatory properties, as reflected by the first three formants, of artificial, non-native, sounds; see Figure 4A.

Next, we analyzed the *attraction index*, in order to determine whether autistic and NT participants differed as to the extent their attempts to reproduce an artificial target were closer, on the F1–F3 acoustic space, to the mean F1-F3 values of a vowel of their own vocalic space than to those of this target. The addition of the Group factor to a multilevel logistic model, with by-item random intercepts, significantly improved the model fit ($\chi_2(1) = 314.95$; $P < 0.001$), the attraction index being significantly higher in the ASD group ($\beta = 0.69$, $se = 0.04$). This effect remained significant when median F0 has

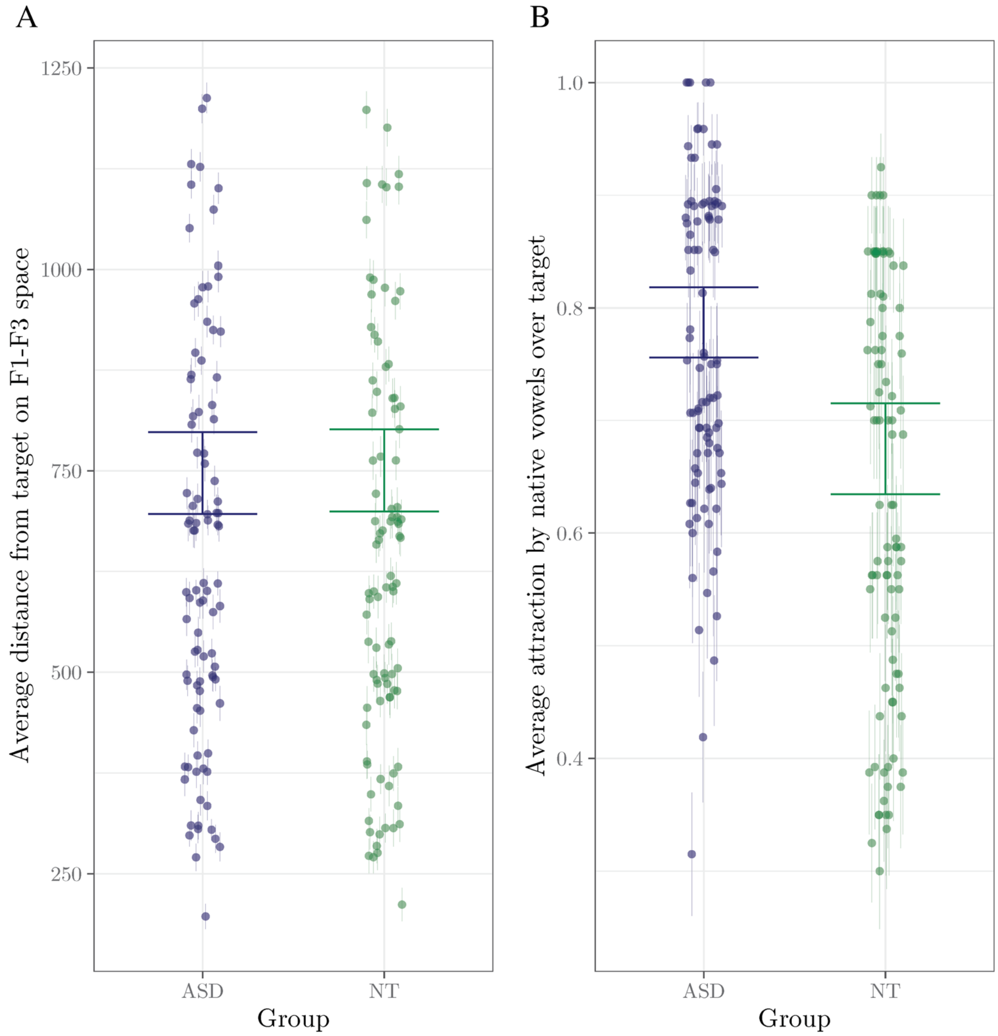

Figure 4. Artificial vowel imitation. (**A**) Average distance from target on F1–F3 space; (**B**) Average attraction index. Points stand for mean values per target and vertical bars for standard errors; superimposed error bars represent fitted 95% confidence intervals.

been added to the regression ($\beta = 0.58$, $se = 0.04$; $p < .001$). That is, albeit autistic participants were not significantly lower in their performance at reproducing artificial vowels (as indicated by the absence of a robust group difference in the distance from the target), they were more attracted by the acoustic properties of their own vocalic space than NT participants; see Figure 4B.

Finally, we analyzed the *dispersion index*, which reflected articulatory variation across different attempts at the production of the same target. The addition of the Group factor improved the model fit for the *dispersion index* ($\chi_2(1) = 117.13$; $P < 0.001$), with the dispersion index being much higher in the NT than in the ASD group ($\beta = 12.68$, $se = 1.17$). This effect remained significant, and, in fact, slightly increased, once the median F0 was introduced within the regression ($\beta = 14.3$, $se = 1.12$; $P < 0.001$). In sum, in line with what has been found for native vowels, both in this study and in Kissine and Geelhand [2019], autistic participants exhibited less variation across different attempts to reach a given target; see Figure 3C.

Discussion

The first objective of this paper was to determine whether autistic individuals exhibit an increased articulatory stability using a controlled task of vowel production. The results of our first task replicate those that Kissine and Geelhand [2019] reported on natural speech: autistic participants display less variation in their realization of the vowels of their native language, as reflected by the first three formants. Furthermore, as in Kissine and Geelhand [2019], our autistic participants displayed an overall higher F0 than NT participants, but a narrower F0 range—that is a more stable F0 during vowel production.

As mentioned in the Introduction, the data analyzed by Kissine and Geelhand [2019] were drawn from the production of a narrative and from a semi-structured conversation. Both speech samples were elicited in the context of the administration of the ADOS, with an unfamiliar experimenter, which raises the possibility that greater articulatory stability in autistic participants was due to a difficulty in adopting a more informal speech register. Our Task 1, however, was completely devoid of any socio-pragmatic constraint. Therefore, the lower dispersion index and the lower F0 range that characterized the production of autistic participants are unlikely to be caused by socio-pragmatic, register selection problems. The fact that autistic participants also displayed less variation in Task 2, that is, in their attempts at imitating non-native artificial vowels constitutes a further indication that low phonetic flexibility is a feature inherent in their speech delivery style. Autistic individuals are often perceived as speaking with an atypical—monotonic or machine-like—tone of voice. While such subjective, qualitative reports are very frequent, to date, no clear objectivization of these impressions is available in the literature. Most studies devoted to the acoustic properties of speech in ASD focused on prosody. From this literature, higher variation in F0 emerges as the only acoustic property that robustly distinguishes autistic participants from neuro-typicals [see Fusaroli et al., 2017, for a meta-analysis]. Such higher variability in F0 seems inconsistent with widespread impressions that the speech of autistic individuals is, on the contrary, rather monotonic. However, most of the studies that report a higher F0 variation in autistic participants groups use naming, reading or narrative production tasks that autistic participants may find challenging [Bone et al., 2015; Bonneh et al., 2011; Diehl et al., 2009; Green & Tobin, 2009; Grossman et al., 2013]. Furthermore, F0 measures are usually collected at a suprasegmental word or utterance level, which may obliterate the low phonetic flexibility uncovered by Kissine and Geelhand [2019] and in this paper. The results of the present paper confirm that research on speech in ASD can benefit from shifting the focus of experimental studies from global prosodic contours to finer-grained segmental properties. Another promising line of research is to explore the extent to which articulatory stability at the phonemic level is related to impressions of an atypical tone of voice. In the introduction, we also discussed two explanatory hypotheses for a greater articulatory stability in ASD. The first line of explanation could be that phonetic inflexibility in ASD is associated with an enhanced capacity to reach precise articulatory targets. The second explanation is that in autistic speakers phonological (gestural) representations are coupled with sensory-motor commands in a less flexible way than in NT individuals. Contrary to the first hypothesis, however, in our second task, autistic participants did not display better performance than NT participants in imitating artificial non-native vowels. In fact, autistic participants' performance tended to be worse in the second task, but the group effect disappeared once the median F0 was controlled for. Given that autistic participants had an overall higher F0, this latter result suggests that matching the stimulus F0 contour represented a further source of difficulty for our autistic participants. Interestingly, though, relative to NT participants, autistic participants' attempts at reproducing non-native vowels were both less variable and more heavily attracted by the characteristics of their own vocalic space. In other words, not only do autistic adults produce vowels and vowel-like sounds in a more stable way than NT speakers, the articulatory properties of their vocalic space also exert a stronger influence on the production of non-native sounds. Note that while it is likely that autistic individuals display an overall reduced flexibility in the realization of phonetic targets, it remains also possible that, within the boundaries of their native phonetic repertoires, autistic adults do display an atypically precise realization of phonetic gestures.

The group differences reported above may reflect intriguing and important characteristics of phonological systems in autism, which, to our minds, deserve to be explored in future research. For instance, it would be interesting to explore lack of phonetic flexibility in relation with early-onset lack of attention to social, communicative cues, amply documented in young autistic children [Jones & Klin, 2013; Zwaigenbaum et al., 2015]. The acquisition of one's native tongue phonological system is underpinned by statistical learning mechanisms, based on distributional frequencies within the speech input. However, in typical development, statistical learning is also heavily bootstrapped by attending to speakers communicative intentions [Kuhl et al., 2008; Kuhl, Tsao, & Liu, 2003; Yeung & Werker, 2009]. Due to poor social orientation and impaired joint attention skills, the extent to which most autistic children can benefit from such socio-communicative cues is very limited. It is therefore possible that phonological development in these children is based on a limited number of prototypes, yielding an atypical phonological acquisition pattern, within which phonological categories become more firmly coupled with a precise sequence of sensory-motor commands than in typical development [Kissine, 2021]. There is also evidence that subtle motor dysfunctions in autism may impact language acquisition [McCleery, Elliott, Sampanis, & Stefanidou, 2013; Stone & Yoder, 2001]. Difficulties in coordinating fine articulatory gestures could also yield a rigid articulatory phonological system. Further studies on child populations, including longitudinal ones, are clearly called for to further explore these ideas.

The sizes of the group effects on dispersion on the F1–F3 spaces in our two tasks, as well as on median F0 and

F0 range in Task 1, are comparable to, and even greater than those in Kissine and Geelhand [2019]. An important difference between our study and Kissine and Geelhand [2019], which is also a clear limitation of the results reported above, is that our participants were only male. Collecting audio recordings only from males yielded homogenous acoustic data. This could explain why the size of our effects was greater than those reported in Kissine and Geelhand [2019], whose sample included females participants. Yet, there is growing evidence that autism in females may be associated with different behavioral characteristics from males [Backer van Ommeren, Koot, Scheeren, & Begeer, 2016; Mandy et al., 2012]. An important next step for future studies is to replicate this study with autistic female participants.

Finally, our tasks differed in format: in Task 1, native vowel production was prompted by written presentation of French graphemes, while Task 2 consisted in imitating auditory stimuli. However, given that in Task 1 only single graphemes or simple and frequent grapheme combinations were presented, this reading component is unlikely to have caused a significant cognitive load for our participants, whose verbal and non-verbal IQs were within typical ranges. Furthermore, the pattern of the group effects is remarkably similar across the two tasks.

Acknowledgments

We thank two anonymous reviewers for their comments on a previous version of this paper. We are very grateful to all our participants who generously gave their time and patience to this study. We are grateful to the authors of Delvaux et al. [2014] for sharing their protocol (including audio materials). We also thank Laura Capouet for her help in data acquisition. Philippine Geelhand is a F.R.S.-FNRS post-doctoral researcher and Mikhail Kissine is a 2019-2022 Francqui Research Professor.

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Backer van Ommeren, T., Koot, H. M., Scheeren, A. M., & Begeer, S. (2016). Sex differences in the reciprocal behaviour of children with autism. Autism, 21(6), 795–803. https://doi.org/10.1177/1362361316669622

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). Springer: Boston.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. Journal of Autism and Developmental Disorders, 34(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31(1), 5–17.

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. [Computer program]. http://www.praat.org/.

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

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.

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., … Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not asperger syndrome. Neuropsychologia, 48(9), 2465–2475.

Boyd, L. E., Rangel, A., Tomimbang, H., Conejo-Toledo, A., Patel, K., Tentori, M., & Hayes, G. R. (2016). Saywat: Augmenting face-to-face conversations for adults with autism, In *Proceedings of the 2016 CHI conference on human factors in computing systems*, San Jose, California, USA. Association for Computing Machinery.

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.

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in Autism Spectrum Disorder: Indirect requests versus irony. Journal of Autism and Developmental Disorders, 48(9), 2938–2952.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2014). Phonetic compliance: A proof-of-concept study. Frontiers in Psychology, 5, 1375.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. Brain and Language, 106(2), 144–152.

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.
Geelhand, P., Deliens, G., Papastamou, F., & Kissine, M. (2020). Narrative production in autistic adults: A systematic analysis of the microstructure, macrostructure and internal state language. Journal of Pragmatics, 164, 57–81.
Geurts, B., Kissine, M., & van Tiel, B. (2020). Pragmatic reasoning in autism. In K. Moranyi & R. Byrne (Eds.), Thinking, reasoning and decision making in autism (pp. 113–134). London: Routledge.
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. https://doi.org/10.1044/1092-4388(2012/12-0067)
Happé, F., & Frith, U. (2006). The Weak Coherence account: Detail-focused cognitive style in Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 36(1), 5–25.
Hobza, V., Hamrik, Z., Bucksch, J., & De Clercq, B. (2017). The family affluence scale as an indicator for socioeconomic status: Validation on regional income differences in The Czech Republic. International Journal of Environmental Research and Public Health, 14(12), 1540.
Johnson, K. (2012). Acoustic and auditory phonetics. Oxford: Wiley.
Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature, 504, 427–431.
Karlsson, F., & van Doorn, J. (2012). Vowel formant dispersion as a measure of articulation proficiency. Journal of the Acoustical Society of America, 132(4), 2633–2641.
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, fourth edition (pp. 230–262). Hoboken: Wiley.
Kissine, M. (2021). Autism, constructionism and nativism. Language. http://homepages.ulb.ac.be/~mkissine/language_perspectives.pdf.
Kissine, M., & Geelhand, P. (2019). Acoustic evidence for increased articulatory stability in the speech of adults with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 49, 2572–2580.
Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67(3), 971–995.
Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (nlm-e). Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1493), 979–1000.
Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences, 100(15), 9096–9101.
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.
Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. Journal of Autism and Developmental Disorders, 42(7), 1304–1313.
McCleery, J., Elliott, N., Sampanis, D., & Stefanidou, C. (2013). Motor development and motor resonance difficulties in autism: Relevance to early intervention for language and communication skills. Frontiers in Integrative Neuroscience, 7, 30.
Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. Journal of Autism and Developmental Disorders, 36(1), 27–43.
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., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). 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), 861–869.
Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A bayesian explanation of autistic perception. Trends in Cognitive Sciences, 16, 504–510.
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, Austria: R Foundation for Statistical Computing.
Scheerer, N. E., Shafai, F., Stevenson, R. A., & Iarocci, G. (2020). Affective prosody perception and the relation to social competence in autistic and typically developing children. Journal of Abnormal Child Psychology, 48, 965–975.
Shriberg, L. D., & Widder, C. J. (1990). Speech and prosody characteristics of adults with mental retardation. Journal of Speech, Language, and Hearing Research, 33(4), 627–653. https://doi.org/10.1044/jshr.3304.627
Stone, W. L., & Yoder, P. J. (2001). Predicting spoken language level in children with Autism Spectrum Disorders. Autism, 5 (4), 341–361.
Torsheim, T., Cavallo, F., Levin, K. A., Schnohr, C., Mazur, J., Niclasen, B., … Group, t. F. D. S. (2016). Psychometric validation of the revised family affluence scale: A latent variable approach. Child Indicators Research, 9(3), 771–784.
Wechsler, D. (2008). Wechsler adult intelligence scale–fourth edition (WAIS–IV) (22). San Antonio, TX. NCS Pearson.

Wehrle, S., Cangemi, F., Hanekamp, H., Vogeley, K., & Grice, M. (2020). Assessing the intonation style of speakers with Autism Spectrum Disorder. In *Speech prosody conference, Tokyo*.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. Cognition, 113(2), 234–243.

Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., … Wetherby, A. (2015). Early identification of Autism Spectrum Disorder: Recommendations for practice and research. Pediatrics, 136, S10–S40.