Auditory Attention Deployment in Young Adults with Autism Spectrum Disorder

Katherine A. Emmons[1], **Adrian KC Lee**[2,3], **Annette Estes**[1,2], **Stephen Dager**[4], **Eric Larson**[3], **Daniel R. McCloy**[3], **Tanya St. John**[1], **Bonnie K. Lau**[3,5]

[1]UW Autism Center, University of Washington, Seattle, WA, USA

[2]Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, USA

[3]Institute for Learning and Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195, USA

[4]Department of Radiology, University of Washington School of Medicine, Seattle, WA, USA

[5]Department of Otolaryngology—Head and Neck Surgery, University of Washington School of Medicine, Seattle, WA, USA

Abstract

Difficulty listening in noisy environments is a common complaint of individuals with autism spectrum disorder (ASD). However, the mechanisms underlying such auditory processing challenges are unknown. This preliminary study investigated auditory attention deployment in adults with ASD. Participants were instructed to maintain or switch attention between two simultaneous speech streams in three conditions: location (co-located versus $\pm 30°$ separation), voice (same voice versus male–female contrast), and both cues together. Results showed that individuals with ASD can selectively direct attention using location or voice cues, but performance was best when both cues were present. In comparison to neurotypical adults, overall performance was less accurate across all conditions. These findings warrant further investigation into auditory attention deployment differences in individuals with ASD.

Keywords

Speech perception; Auditory attention; Auditory processing; Selective attention; Autism spectrum disorder

Introduction

Communication in everyday life depends crucially on the ability to dynamically shift attention between competing auditory streams. Many of us do so effortlessly, such as switching attention between different speakers at a party. Despite the apparent simplicity, selectively attending to one of several speakers is a complex task, involving perceptual as

---

✉Bonnie K. Lau, blau@uw.edu.

well as cognitive processes (Lee et al., 2014; Shinn-Cunningham et al., 2017). Listeners must first segregate competing voices into separate auditory streams using perceptual cues such as the location and voice characteristics of the target speaker (Carlyon, 2004; Darwin, 1997) and then selectively direct attention on the target speaker while ignoring competing speakers. Whereas auditory attention deployment has been well studied in neurotypical (NT) adults with normal hearing, it is relatively under-studied in individuals with autism spectrum disorder (ASD), who often demonstrate auditory processing differences (reviewed by Haesen et al., 2011; O'Connor, 2012). This study aims to investigate auditory attention deployment in young adults with ASD and NT participants.

ASD is a neurodevelopmental disorder characterized by restrictive and repetitive behaviors, as well as social and communication challenges (American Psychiatric Association, 2013). Although it is not a core feature of ASD, impairments of attention such as difficulty switching attention between tasks, abnormal distribution of attentional resources, and impaired cross-modal attention switching are commonly observed in adults and children with ASD (Allen & Courchesne, 2001; Reed & McCarthy, 2012). For example, school-aged children with ASD show greater difficulty switching attention between visual tasks compared to typically developing children, and particular difficulty switching attention between visual and auditory modalities (Reed & McCarthy, 2012). In the visual domain, individuals with ASD may demonstrate abnormally narrow fields of attention, as well as difficulty disengaging attention from a visual target (Allen & Courchesne, 2001).

In the auditory domain, adults and children with ASD often demonstrate processing differences such as hyper- and hyposensitivity to sound, atypical orientation to auditory information, difficulty listening under noisy conditions, and inefficient auditory stream segregation (reviewed by Haesen et al., 2011; O'Connor, 2012). Past studies also suggest that adults with ASD have poor sound source localization: they are slower and less accurate in localizing noise bursts presented from speakers in the azimuthal plane compared to control participants (Teder-Sälejärvi et al., 2005).

Auditory processing differences have also been demonstrated specifically for speech sounds in ASD (Ceponiene et al., 2003; Jones et al., 2020; Kuhl et al., 2005; O'Connor, 2012; Otto-Meyer et al., 2018). Toddlers with ASD may prefer non-speech sounds whereas typically developing (TD) children show a preference for infant-directed speech (Kuhl et al., 2005). Differences in auditory brainstem responses to speech sounds have also been reported in toddlers with ASD, suggesting inefficient processing of speech in comparison to their TD peers (Jones et al., 2020). Similar results have been observed in school-aged children with ASD: compared to TD children, children with ASD have less stable frequency-following responses to certain speech sounds, which may contribute to deficits in language and communication (Otto-Meyer et al., 2018).

Selective auditory attention in NT adults comes with a cost: listeners typically perform worse when asked to switch attention between talkers compared to maintaining attention on a single talker (Larson & Lee, 2013a, 2013b; Lawo & Koch, 2015; McCloy et al., 2017, 2018). Responses are slower and less accurate when participants are required to switch attention between competing auditory streams (Larson & Lee, 2013b). Switching attention

between competing auditory streams also bears a greater cognitive load, as observed by increased pupillary responses (McCloy et al. 2017). Relatively few studies on auditory attention switching have focused on adults with ASD. Consequently, despite the broad range of auditory processing differences observed in adults and children with ASD, it is unclear whether adults with ASD demonstrate the same "switch cost" observed in NT adults.

To selectively direct attention to the desired speaker, listeners must first segregate competing voices into distinct auditory streams using a variety of perceptual cues. Binaural cues, or timing and intensity differences between the two ears, play a critical role in segregating competing talkers and help listeners localize where the voice of interest is coming from. For example, if the speaker of interest is standing to a listener's right, their voice will arrive at the right ear faster and also louder than the left ear. The listener's auditory system can use these binaural cues to differentiate the speaker of interest from the other competing talkers. When speakers' voices are coming from the same location, or co-located, listeners are not able to use binaural cues because the voices arrive at the ears at the same time and loudness. To segregate competing voices in this situation, they must rely on voice characteristics such as pitch or loudness of the particular speaker.

NT listeners also benefit when multiple auditory streams come from different locations such that they can be spatially segregated (Bronkhorst, 2015; Brungart & Simpson, 2007; Ericson et al., 2004). Ericson et al. (2004) investigated the role of binaural cues on listener performance on a multi-talker listening task with one, two, or three competing speakers. They found that with one competing speaker, spatial segregation of the target and competing voices improved listener performance by approximately 25%; with two or three competing speakers, spatial segregation nearly doubled the percentage of correct responses (Ericson et al., 2004). Similarly, listeners perform better on a multi-talker speech perception tasks when competing talkers are spaced far apart, as opposed to close together (Brungart & Simpson, 2007).

Prior research in NT adults also suggests listeners benefit when simultaneous auditory streams are composed of voices of different sexes (reviewed by Bronkhorst, 2015). For example, Brungart (2001) studied listener intelligibility of a target phrase masked by a competing phrase and found that listener performance was best when the target and the competing voices were different sexes. Likewise, Darwin et al. (2003) studied the effect of fundamental frequency and vocal-tract length changes on listeners ability to attend to one of two simultaneous sentences and found that listener performance increased dramatically when voices comprised speakers of different sexes. Despite the evidence regarding auditory attention deployment in NT adults, little is known about how auditory attention deployment differs in adults with ASD, and how these differences may contribute to social and communication challenges.

When multiple perceptual cues are available, such as both location and voice characteristics, listeners may find the most benefit (reviewed by Bronkhorst, 2015). For example, listeners are more accurate in identifying simultaneously presented vowel sounds when they differ in both pitch (F0) and location, compared to either cue on its own (Du et al., 2011). A priori knowledge of the target speaker's voice characteristics and location may also

improve listener performance on selective listening tasks (Shinn-Cunningham & Best, 2008), particularly when there are two or more competing speakers (Ericson et al., 2004). Relatively few studies have looked specifically at how adults with ASD use perceptual cues, including location and voice cues, to segregate simultaneous auditory streams and how that may differ from the patterns observed in NT adults.

This preliminary study was designed to investigate auditory attention deployment in young adults with ASD. Participants were presented with a selective listening task where they were asked to maintain or switch attention between one of two simultaneous auditory streams and repeat back two words as instructed. The spatial location of the two speech streams (co-located vs. spatially separated) as well as voice characteristics of the talkers (same voice versus male–female contrast) were also manipulated. Three research questions were investigated. First, do individuals with ASD perform worse when asked to switch attention between auditory streams? We hypothesized that participants with ASD would perform worse on switch attention compared to maintain attention trials, and thus, show a "switch cost" like NT adults. Second, can individuals with ASD make use of spatial location and talker voice cues to improve their ability to selectively listen to simultaneous speech streams? We hypothesized that they would perform best when provided with both location and voice cues. Third, do individuals with ASD show greater difficulty maintaining or switching attention between speech streams than age- and sex-matched comparison participants? We hypothesized that individuals with ASD would demonstrate greater difficulty maintaining and switching auditory attention across all conditions than comparison group participants.

Method

Participants

Twenty-four adults (aged 21–23 years) participated in the study. Our sample size was determined based on an a priori power analysis to detect group differences in a repeated measures analysis of variance (ANOVA) with 2 groups, 6 measurements (2 attention conditions by 3 cue conditions), and a correlation among repeated measures of 0.5. A sample size of 11 subjects per group was determined as sufficient to detect a group difference with a medium effect size of 0.5 with a power (1-β) of 0.8 and probability (α) of 0.05. Twelve participants with ASD were recruited from a larger longitudinal study conducted at the University of Washington. The original cohort consisted of 72 children diagnosed with ASD between the ages of 3 and 4 years. Diagnoses of Autistic Disorder, Asperger's Disorder, or Pervasive Developmental Disorder—Not Otherwise Specified, were made according to criteria from the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; American Psychiatric Association, 1994) by a licensed clinical psychologist or supervised graduate student using: (1) the Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994), (2) the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 1989), (3) medical and family history, (4) cognitive test scores, and (5) clinical observation and judgment (see Dawson et al., 2004 for details). These participants were tested again at age 6, 9, and 13–15 years. Forty-six participants from the original cohort were re-contacted and invited to participate in this

current study. The remaining 26 were not contacted because they were already recruited for another study. Of the 46 participants contacted, 4 moved out of state, 2 were not interested in participating, 1 could not be scheduled, 24 did not respond to phone calls or emails, 2 did not meet our eligibility criteria of being able to speak in 3-word phrases, and 1 failed the audiometric screening. Twelve participants with ASD from the original cohort were enrolled in the study and tested at age 21–23 years of age. Twelve typically developing participants aged 21–23 years who did not participate in the original study were also newly recruited as age- and sex-matched comparison participants. Comparison participants reported no history of cognitive, developmental, or other health concerns.

Autism Diagnostic Observation Schedule—Second Edition (ADOS-2; Lord et al., 2012) and Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011) scores were obtained for all 24 participants. One NT participant failed to complete the selective listening task because they were not able to tolerate the testing space (i.e., the participant reported having claustrophobia). Selective listening task scores from one ASD participant were omitted from group analyses due to poor data quality (audio recording was unintelligible). This resulted in a final sample size of 22 participants (n = 11 ASD; n = 11 NT) (Table 1).

Procedures

The following measures were obtained over the course of several visits and as part of a larger study that included additional neurophysiological and behavioral measures. These results have not yet been published. Written informed consent was obtained from all participants in accordance with protocols reviewed and approved by the Institutional Review Board at the University of Washington.

ADOS-2

The ADOS-2 (Lord et al., 2012), a measure of autism symptom severity, was administered to all participants at the time of testing to confirm group inclusion (ASD vs. Comparison). All participants in the ASD group received a classification of autism or autism spectrum based on their performance on the ADOS-2. No participants in the comparison group received a classification of autism or autism spectrum based on their performance on the ADOS-2.

WASI-II

The WASI-II (Wechsler, 2011), a measure of intellectual ability, was administered to all participants at time of testing. The WASI-II Full Scale Intelligence Quotient (FSIQ-4), computed from the subtests: block design, vocabulary, matrix reasoning, and similarities, was obtained from each participant and provides an overall estimate of their general level of intellectual functioning.

Audiological Screening

There is a higher incidence of hearing loss in individuals with ASD (Klin, 1993; Rosenhall et al., 1999; Szymanski et al., 2012). In individuals with comorbid ASD and hearing loss, it would be difficult to differentiate whether it is hearing loss or auditory attention

deployment that affects performance on our selective attention task. Thus, normal hearing was an inclusion criterion to participate in this study. To ensure clinically normal hearing thresholds, all participants were required to pass an audiometric screen (20 dB hearing level at octave frequencies between 250 and 8000 Hz), a distortion product otoacoustic emission (DPOAE) screen, and an auditory brainstem response (ABR) screen.

Selective Listening Task

Participants completed the selective listening task while undergoing Magneto- and Electroencephalography (M-EEG) recording. Each trial began with an instruction spoken by the target speaker (male or female; see trial configuration illustrated in Fig. 1). The instruction was either "hold attention" or "switch attention", which indicated whether to maintain attention on the target speaker for the duration of the trial, or switch attention to the other speaker midway through the trial. 1500 ms after the instruction onset, two simultaneous auditory streams began: the target stream to be attended and the distractor stream to be ignored. Each trial belonged to one of three cue conditions: spatial location, talker voice, or both cues presented together. For trials in the spatial location condition, the two streams were the same male or female voice spatialized to left and right sides at ±30° azimuth (Fig. 2a). For trials in the talker voice condition, the two streams were a male and female voice co-located at either +30° or −30° azimuth (Fig. 2b). For trials with both cues, the two streams were a male and female voice, one spatialized to the left side and the other to the right side at ±30° (Fig. 2c). The participant's task was to repeat back two compound words (e.g., "headphones, football") from the target speaker at the end of each trial. On maintain attention trials, that would be the two words spoken by the target speaker. On switch attention trials, that would be the first word spoken by the target speaker and the second word spoken by the other speaker.

There were 8 blocks of 24 trials each, for a total of 192 trials. The sex and location of the target speaker was fixed within each block of trials. The two attention conditions (maintain, switch) and three cue conditions (spatial location, talker voice, both) were counterbalanced and intermixed within each block. Block order was randomly assigned for each participant.

On a separate day, participants completed two training blocks of the selective listening task which consisted of 24 trials each. During the training period, participants were told the correct answer after each trial. On the day of data acquisition, participants were offered one block of trials as a reminder, with no data taken. Participants were not given feedback during testing.

Auditory Stimuli

To generate the auditory stimuli two talkers (one male and one female) were fitted with head-mounted close-talking microphones. The four target words (football, footprint, headlight, headphones) were recorded a minimum of six times per talker. The clearest exemplar of each syllable in the target words was extracted from each talker's recording using Praat and was resynthesized to monotonize the pitch to the mean pitch of that talker. Duration was also adjusted (separately for initial and final syllables) using Praat's implementation of the PSOLA algorithm, to equate duration across all stimuli from either

talker. Stimuli were resampled in the frequency domain using the SciPy library (Virtanen et al., 2020) from 44,100 to 24,414 Hz for compatibility with stimulus delivery hardware. Two streams of two words were generated for each trial, for a total of 192 unique trials. Auditory stimuli were presented using sound-isolating tubal insertion earphones (Etymotic ER-2). Auditory stimuli were presented at 65 dB sound pressure level (SPL) against a white-noise background with pi-interaural-phase at 45 dB SPL. Speakers' voices were spatialized to left and right sides using convolution with CIPIC head-related transfer functions (Algazi et al., 2001).

Participant responses to the auditory stimuli were coded as "correct" if the subject successfully repeated back the two words spoken by the target speaker in the correct order. For example, in the maintain trial depicted in Fig. 1, the correct answer is "headphones, football." If the participant said "headphones, football" during the response period, then their response was coded as correct and they were awarded one point. Participants were not awarded partial credit. For example, a response of "headphones, headlight" was coded as incorrect and the participant received zero points. Unintelligible responses were coded as incorrect.

Results

Attentional Demand

To investigate whether participants in the ASD group performed worse when asked to switch attention between the two auditory streams, paired samples t-tests comparing scores on maintain-attention versus switch-attention trials were conducted within the ASD group. Average performance on the maintain attention trials was 71% correct ($SD = 14.4$) while average performance on the switch-attention trials was 60.5% correct ($SD = 20.6$). The results of the paired samples t-test were not significant (Fig. 3, light blue vs. dark blue, $t(10) = 1.67$, p = 0.127, *Cohen's d* = 0.50) indicating that the switch-attention trials were no more difficult than the maintain-attention trials for the ASD group.

Cue Condition

To investigate whether participants in the ASD group were able to use location and voice cues to selectively listen to simultaneous speech streams, scores in each attention condition were submitted to a one-way analysis of variance. Maintain-attention scores were submitted to a repeated-measures ANOVA with within-subjects factor Cue Condition. A significant main effect of Cue Condition on performance on maintain-attention trials was observed (Fig. 4a *left*; $F(2, 20) = 9.70$, p = 0.001, $n_p^2 = 0.492$). Bonferroni adjusted post hoc pairwise comparisons showed better performance on maintain-attention trials with both cues compared to trials with just a voice cue (Fig. 4a *left*, light blue vs. dark blue bar; post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p = 0.012) or just a location cue (Fig. 4a *left*, light blue vs. green bar; post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p < .001).

Next, switch-attention task scores for individuals with ASD were submitted to a repeated-measures ANOVA with within-subjects factor Cue Condition. A significant main effect of Cue Condition on performance on switch-attention trials was also observed (Fig. 4b *left*;

$F(2, 20) = 7.42$, p = 0.004, $n_p^2 = 0.426$). Bonferroni adjusted post hoc pairwise comparisons also showed better performance on switch-attention trials with both cues compared to trials with just a voice cue (Fig. 4b *left*, light blue vs. dark blue bar; post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p = 0.015) or just a location cue (Fig. 4b *left*, light blue vs. green bar; post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p = 0.007), as in the maintain-attention trials.

Overall Task Performance Comparison

NT adults achieved a higher percent correct across all conditions (Fig. 4 *right*) compared with the ASD group. Overall, the NT group scored 90.7% ($SD = 11.3$) correct on maintain trials and 85.2% correct ($SD = 13.3$) on switch trials compared to 71.0% ($SD = 14.4$) and 60.5% ($SD = 20.6$) in the ASD group. To investigate how overall task performance differed in the ASD group compared to the comparison group, task scores from both groups were submitted to an ANOVA. A repeated-measures mixed ANOVA on task scores with between-subjects factor Group (ASD, NT), and within-subjects factors Attention Condition (Maintain, Switch) and Cue Condition (Location, Voice, Both) revealed significant main effects of group ($F(1, 20) = 15.45$, p = 0.001, $n_p^2 = 0.436$), Attention Condition ($F(1, 20) = 5.96$, p = 0.024, $n_p^2 = 0.229$), and Cue Condition ($F(2, 40) = 15.50$, p < 0.001, $n_p^2 = 0.437$). These results show that overall, we observed better performance in the NT group compared to the ASD group (ASD M = 65.77, $SE = 4.00$;) (ASD M = 65.77, $SE = 4.00$; NT M = 87.97, $SE = 4.00$), better performance on maintain-attention compared to switch-attention trials (maintain $M = 80.87$, $SE = 2.76$; switch $M = 72.87$, $SE = 3.70$) and an effect of Cue Condition on performance.

Bonferroni adjusted post hoc pairwise comparisons on Cue Condition revealed better performance on trials with both cues compared to trials with just a voice cue (post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p < 0.001) or just a location cue (post hoc t-test, Bonferroni adjusted $\alpha = 0.017$, p < 0.001). We also observed a significant Cue Condition by Group interaction ($F(2, 40) = 3.31$, p = 0.047, $n_p^2 = 0.142$), indicating that the effect of Cue Condition differed between the ASD group and the NT group. That is, participants in the ASD group performed better on trials with both spatial and voice cues while no difference was observed across cue conditions in the NT group. No other significant interactions were observed (Attention Condition by Group: p = 0.453; Attention Condition by Cue Condition: p = 0.203; Attention Condition by Cue Condition by Group: p = 0.964).

Intellectual Ability and Task Performance

As the two groups differed significantly in intellectual ability ($t(20) = -3.29$, p = 0.004, *Cohen's d* = 1.40) we performed a follow-up exploratory analysis on the effect of intellectual ability on task performance. A linear regression was conducted to assess the relationship between WASI-II FSIQ-4 and overall performance on the selective attention task. WASI-II FSIQ-4 was not a significant predictor of overall performance on the selective attention task ($\beta = 0.483$, $t = 1.83$, p = 0.082, $r^2 = 0.144$) suggesting no clear correlation between WASI-II FSIQ-4 score and performance on the selective attention task in these participants. To further investigate the relationship within each group, separate linear regressions were conducted for participants in the ASD and NT groups to assess the

relationship between WASI-II FSIQ-4 and overall task performance. WASI-II FSIQ-4 was not a significant predictor of overall task performance in the ASD group ($\beta = -0.156$, $t = -0.472$, p = 0.648, $r^2 = 0.024$) nor the NT group ($\beta = 0.603$, $t = 1.06$, p = 0.317, $r^2 = 0.111$).

Qualitative inspection of the data also supports these findings. For example, the two participants with the lowest FSIQ-4 scores (76 and 81) performed vastly differently on the task, with one participant averaging about 65% correct and the other participant averaging about 93% correct across conditions. Moreover, the participant with the lowest average task performance (about 43% correct) had a WASI-II FSIQ-4 score within 1 standard deviation of the mean (WASI-II FSIQ-4 = 96).

Discussion

The present study was designed as a preliminary investigation of auditory attention deployment abilities in young adults with ASD using a selective listening task. We observed three main findings in this study. First, no difference in performance on maintain-attention compared to switch-attention trials was present in the ASD group, contrary to what we predicted. While previous studies reported a "switch cost" (i.e., worse performance on switch attention trials) of auditory attention in NT adults, this difference was not found in our cohort of individuals with ASD. Second, participants with ASD performed best on trials with both location and voice cues compared to just one cue, suggesting that they were able to use both cues to improve performance on this selective attention task. Finally, we observed overall worse task performance in the ASD group compared to the NT group, with lower scores across all experimental conditions, suggesting difficulty deploying auditory attention in the individuals with ASD. Results from our study provide preliminary evidence that although participants with ASD were able to use pre-trial instructions as well as spatial location and talker voice cues to selectively direct auditory attention, they demonstrated greater difficulty on these tasks than NT participants. This finding warrants further studies with larger sample sizes into auditory attention deployment in ASD.

Individuals with ASD not only showed difficulty switching attention between competing auditory streams, but they also showed difficulty maintaining attention on one of two simultaneous auditory streams. This result suggests broad difficulties with selective auditory attention in ASD, consistent with past research on general auditory processing differences (reviewed by Haesen et al., 2011; O'Connor, 2012). While much of the research on auditory processing challenges in ASD has focused on children and adolescents, results from our study suggest difficulties with auditory attention deployment may persist into early adulthood. Furthermore, participants with ASD in our study all had normal hearing thresholds. Therefore, difficulties listening under noisy conditions may result from inefficient or abnormal integration of sensory information, rather than an inability to perceive sound.

One initially surprising aspect of our findings was that individuals with ASD did not demonstrate a "switch cost" when asked to switch attention between the two streams. Prior studies such as Larson and Lee (2013b) and McCloy et al. (2017) showed that NT adults performed worse on switch-attention compared to maintain-attention trials. However,

closer inspection of the data revealed that ASD participants also performed worse than NT participants on the maintain-attention trials, suggesting that both the maintain-attention and switch-attention conditions were equally challenging, and thus, did not demonstrate a "switch cost". One possible explanation for this finding is that the ASD group had difficulties with segregating the auditory streams themselves, regardless of whether they were required to maintain or switch attention on a given stream. This explanation would be consistent with Lepistö et al. (2009), who observed inefficient auditory stream segregation in school aged children with Asperger syndrome.

As expected, we observed an additive effect of perceptual cues in the ASD group; participants performed better on trials with both cues compared to trials with just one cue, suggesting that they were able to use both location and voice cues. This finding is consistent with Du et al. (2011), who observed an additive effect of perceptual cues in normal hearing, NT adults. One important difference between our study and Du et al. (2011), however, was the auditory stimulus used. In our study, participants had to listen for compound words (e.g., headlight, football, etc.) as opposed to the single vowel sounds used in Du et al. (2011). When listening for a compound word, listeners may be able to identify the word even if they do not fully hear each sound. Therefore, our task is easier than the one used in Du et al. (2011). Nevertheless, individuals with ASD found benefit from having the additional second cue.

Difficulties with auditory stream segregation and selective auditory attention may contribute to the daily life communication challenges experienced by individuals with ASD, such as difficulty listening under noisy, real-world conditions. The results from this preliminary study indicate that individuals with ASD may struggle to maintain attention on just one speaker in the presence of competing speakers, such as at a party or in a classroom. Moreover, our results might suggest that individuals with ASD could benefit from having more perceptual cues available when distinguishing between several competing speakers. For example, this may translate to having better success listening at a party when having their communication partner step away from competing background speakers, especially when everyone shares similar voice characteristics (e.g., same gender voice).

There are several possible mechanisms to explain the preliminary observed differences in auditory attention deployment in adults with ASD, particularly the observation that while NT participants showed no difference in performance on trials with both cues compared to just one cue, ASD participants performed better on trials with both cues. One possibility is that individuals with ASD struggle using location cues to segregate competing auditory streams. In both maintain- and switch-attention conditions, task performance in the ASD group was lowest when speakers came from different spatial locations but comprised voices of the same sex. However, this was also the case for comparison participants. So, difficulty using location cues alone to segregate competing voices seems to be challenging for all participants based on the spatial separation used in this experiment (viz., ±30°), regardless of ASD status. Another possibility is that individuals with ASD struggle using talker voice cues. Compared to the NT group, who performed the same on trials with just a voice cue compared to trials with both cues, participants in the ASD group performed worse on trials with just a voice cue compared to trials with both cues. One alternative explanation may be

that individuals with ASD do not show particular difficulty with any one type of perceptual cue, but rather, that they benefit from having multiple perceptual cues to selectively direct auditory attention. Further research is needed to understand the underlying mechanisms involved in patterns of auditory attention deployment in young adults with ASD. Despite our small sample size, results from our study provide initial evidence that different patterns of auditory attention deployment are observed in young adults with ASD, and that such differences may contribute to social and communication challenges.

Individual differences in task performance is another possible explanation for the observed differences in auditory attention deployment in adults with ASD. Indeed, we see a high degree of individual variability in task performance in the ASD group. Some participants performed very well, on par with comparison participants, whereas others performed quite poorly. It is possible, therefore, that the poor performance of some individuals in the ASD group are responsible for the low group averages. However, we also observed individual variability in the NT group, with some participants performing very well and others performing at a similar level to ASD participants. Hence, it is unlikely that individual variability fully explains the observed differences in task performance in the ASD group and the NT group.

One important limitation to the current study is the small sample size ($n = 11$ ASD; $n = 11$ NT). While a larger sample size would certainly add power to the analyses, this dataset provided the opportunity to generate new hypotheses and draw preliminary conclusions. Future research is needed to replicate this experiment in a larger sample. It should also be noted that the results from our study may not replicate across different age groups; different patterns of auditory attention deployment may be observed in children or older adults with ASD. Future studies should aim to investigate whether the trends observed in our study are consistent across different age groups.

In conclusion, preliminary results from our study suggest individuals with ASD can use pre-trial instructions as well as spatial location and talker voice cues to selectively direct auditory attention, though less accurately than comparison participants. Overall, individuals with ASD performed worse on the selective listening task compared to NT participants, with lower scores across all experimental conditions. Participants with ASD performed better when multiple perceptual cues were available. Our results provide early evidence of different patterns of auditory attention deployment in young adults with ASD. This warrants further studies with larger sample sizes to better characterize the differences between auditory attention deployment in individuals with ASD and NT adults as well as investigate possible mechanisms underlying such differences. Overall, results from our study demonstrate the importance of understanding differences in auditory attention deployment in ASD and how they may contribute to daily life communication challenges.

Acknowledgments

References

Algazi VR, Duda RO, Thompson DM, & Avendano C (2001). The cipic hrtf database. In Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575), 24 Oct 2001 (pp. 99–102). doi:10.1109/ASPAA.2001.969552

Allen G, & Courchesne E (2001). Attention function and dysfunction in autism. Frontiers in Bioscience, 6, D105–119. 10.2741/alien [PubMed: 11171544]

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders. (4th ed.). American Psychiatric Association.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. (5th ed.). American Psychiatric Association.

Bronkhorst AW (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. Attention, Perception, and Psychophysics, 77(5), 1465–1487. 10.3758/s13414-015-0882-9

Brungart DS (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. Journal of the Acoustical Society of America, 109(3), 1101–1109. 10.1121/1.1345696

Brungart DS, & Simpson BD (2007). Cocktail party listening in a dynamic multitalker environment. Perception and Psychophysics, 69(1), 79–91. 10.3758/bf03194455 [PubMed: 17515218]

Carlyon RP (2004). How the brain separates sounds. Trends in Cognitive Sciences, 8(10), 465–471. 10.1016/j.tics.2004.08.008 [PubMed: 15450511]

Ceponiene R, Lepistö T, Shestakova A, Vanhala R, Alku P, Näätänen R, & Yaguchi K (2003). Speech-sound-selective auditory impairment in children with autism: They can perceive but do not attend. Proceedings of the National Academy of Sciences, 100(9), 5567–5572. 10.1073/pnas.0835631100

Darwin CJ (1997). Auditory grouping. Trends in Cognitive Sciences, 1(9), 327–333. 10.1016/S1364-6613(97)01097-8 [PubMed: 21223942]

Darwin CJ, Brungart DS, & Simpson BD (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. Journal of the Acoustical Society of America, 114(5), 2913–2922. 10.1121/1.1616924

Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A, & Liaw J (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. Developmental Psychology, 40(2), 271–283. 10.1037/0012-1649.40.2.271 [PubMed: 14979766]

Du Y, He Y, Ross B, Bardouille T, Wu X, Li L, & Alain C (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. Cerebral Cortex, 21(3), 698–707. 10.1093/cercor/bhq136 [PubMed: 20685854]

Ericson MA, Brungart DS, & Simpson BD (2004). Factors that influence intelligibility in multitalker speech displays. The International Journal of Aviation Psychology, 14(3), 313–334. 10.1207/s15327108ijap1403_6

Haesen B, Boets B, & Wagemans J (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. Research in Autism Spectrum Disorders, 5(2), 701–714. 10.1016/j.rasd.2010.11.006

Jones MK, Kraus N, Bonacina S, Nicol T, Otto-Meyer S, & Roberts MY (2020). Auditory processing differences in toddlers with autism spectrum disorder. Journal of Speech, Language, and Hearing Research, 63(5), 1608–1617. 10.1044/2020_JSLHR-19-00061

Klin A (1993). Auditory brainstem responses in autism: Brainstem dysfunction or peripheral hearing loss? Journal of Autism and Developmental Disorders, 23(1), 15–35. 10.1007/BF01066416 [PubMed: 8463195]

Kuhl PK, Coffey-Corina S, Padden D, & Dawson G (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. Developmental Science, 8(1), F1–f12. 10.1111/j.1467-7687.2004.00384.x [PubMed: 15647058]

Larson E, & Lee AKC (2013). The cortical dynamics underlying effective switching of auditory spatial attention. NeuroImage, 64, 365–370. 10.1016/j.neuroimage.2012.09.006 [PubMed: 22974974]

Larson E, & Lee AKC (2013). Influence of preparation time and pitch separation in switching of auditory attention between streams. The Journal of the Acoustical Society of America, 134(2), EL165–EL171. 10.1121/1.4812439 [PubMed: 23927220]

Lawo V, & Koch I (2015). Attention and action: The role of response mappings in auditory attention switching. Journal of Cognitive Psychology, 27(2), 194–206. 10.1080/20445911.2014.995669

Lee AK, Larson E, Maddox RK, & Shinn-Cunningham BG (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. Hearing Research, 307, 111–120. 10.1016/j.heares.2013.06.010 [PubMed: 23850664]

Lepistö T, Kuitunen A, Sussman E, Saalasti S, Jansson-Verkasalo E, Nieminen-von Wendt T, & Kujala T (2009). Auditory stream segregation in children with asperger syndrome. Biological Psychology, 82(3), 301–307. 10.1016/j.biopsycho.2009.09.004 [PubMed: 19751798]

Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. Journal of Autism and Developmental Disorders, 19(2), 185–212. 10.1007/bf02211841 [PubMed: 2745388]

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. 10.1007/bf02172145 [PubMed: 7814313]

Lord C, Rutter M, DiLavore P, Risi S, Gotham K, & Bishop S (2012). Autism diagnostic observation schedule–(ADOS-2). Western Psychological Corporation.

McCloy DR, Larson E, & Lee AKC (2018). Auditory attention switching with listening difficulty: Behavioral and pupillometric measures. Journal of the Acoustical Society of America, 144(5), 2764. 10.1121/1.5078618

McCloy DR, Lau BK, Larson E, Pratt KAI, & Lee AKC (2017). Pupillometry shows the effort of auditory attention switching. Journal of the Acoustical Society of America, 141(4), 2440. 10.1121/1.4979340

O'Connor K (2012). Auditory processing in autism spectrum disorder: A review. Neuroscience and Biobehavioral Reviews, 36(2), 836–854. 10.1016/j.neubiorev.2011.11.008 [PubMed: 22155284]

Otto-Meyer S, Krizman J, White-Schwoch T, & Kraus N (2018). Children with autism spectrum disorder have unstable neural responses to sound. Experimental Brain Research, 236(3), 733–743. 10.1007/s00221-017-5164-4 [PubMed: 29306985]

Reed P, & McCarthy J (2012). Cross-modal attention-switching is impaired in autism spectrum disorders. Journal of Autism and Developmental Disorders, 42(6), 947–953. 10.1007/s10803-011-1324-8 [PubMed: 21720723]

Rosenhall U, Nordin V, Sandstrom M, Ahlsen G, & Gillberg C (1999). Autism and hearing loss. Journal of Autism and Developmental Disorders, 29(5), 349–357. 10.1023/a:1023022709710

Shinn-Cunningham B, Best V, & Lee AKC (2017). Auditory object formation and selection. In Middlebrooks JC, Simon JZ, Popper AN, & Fay RR (Eds.), The auditory system at the cocktail party. Springer.

Shinn-Cunningham BG, & Best V (2008). Selective attention in normal and impaired hearing. Trends in Amplification, 12(4), 283–299. 10.1177/1084713808325306 [PubMed: 18974202]

Szymanski CA, Brice PJ, Lam KH, & Hotto SA (2012). Deaf children with autism spectrum disorders. Journal of Autism and Developmental Disorders, 42(10), 2027–2037. 10.1007/s10803-012-1452-9 [PubMed: 22290585]

Teder-Sälejärvi WA, Pierce KL, Courchesne E, & Hillyard SA (2005). Auditory spatial localization and attention deficits in autistic adults. Cognitive Brain Research, 23(2–3), 221–234. 10.1016/j.cogbrainres.2004.10.021 [PubMed: 15820630]

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Joshua Wilson K, Millman J, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson Eric, … van Mulbregt P (2020). Scipy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. 10.1038/s41592-019-0686-2 [PubMed: 32015543]
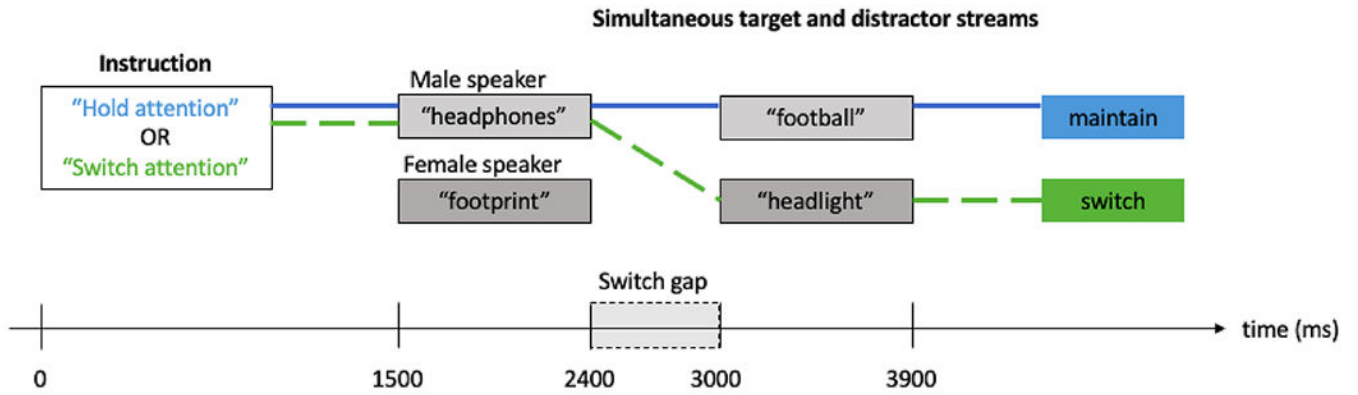
Wechsler D (2011). Wechsler abbreviated scale of intelligence–second edition. Minneapolis: NCS Pearson.

Simultaneous target and distractor streams

Fig. 1.

Trial configuration. Trials began with an auditory instruction (spoken phrase, "hold attention" or "switch attention") spoken by the target speaker (male or female voice from the left side or the right side). 1500 ms after the cue onset, two simultaneous auditory streams began, with a 600 ms gap between the first and second words. In the maintain trial depicted above (solid line), subjects would hear "hold attention" in a male voice from the left or right side, attend to the male voice throughout the trial and repeat back, "headphones, football."
In the switch trial depicted above (dashed line), subjects would hear "switch attention" in a male voice from the left or right, attend to the male voice for the first word, then switch attention to the female voice for the second word and repeat back, "headphones, headlight."
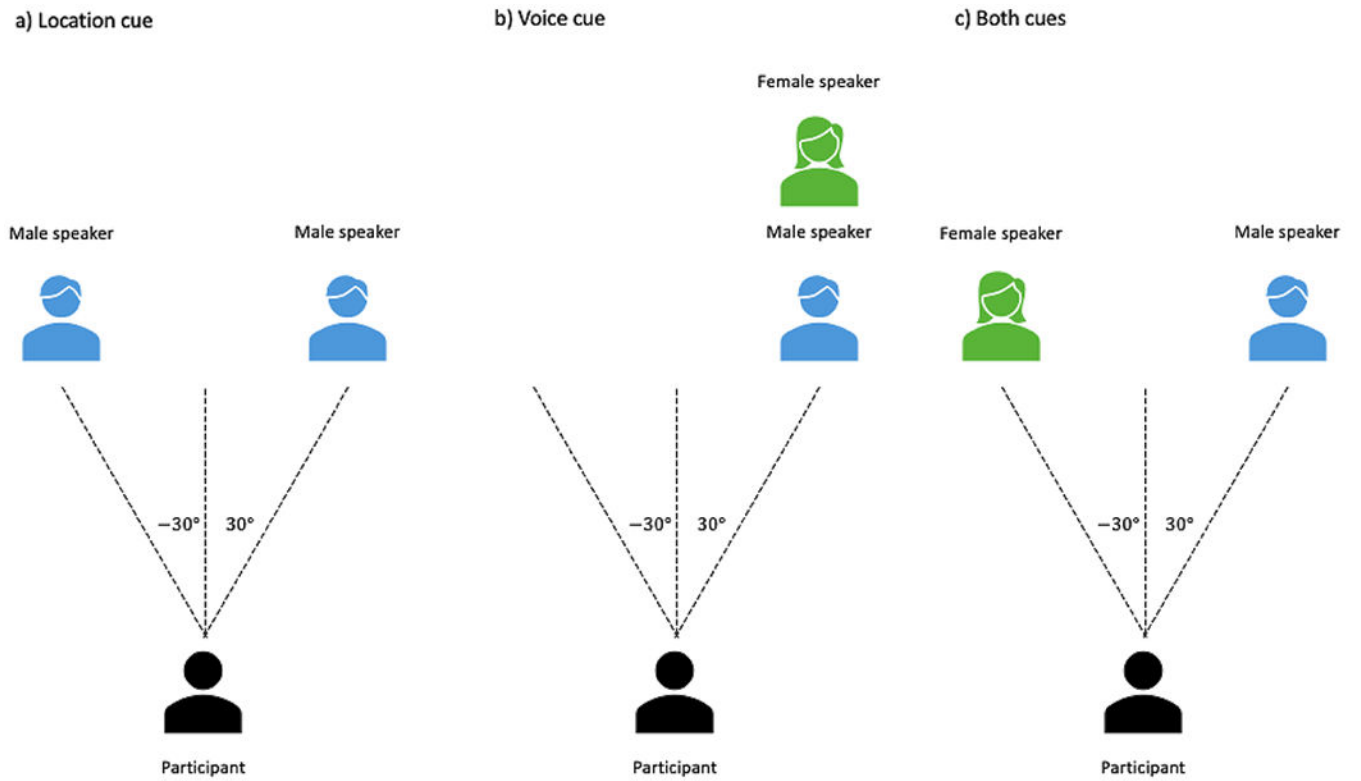
Fig. 2.
Schematic diagram of cue conditions. **a** location cue: same male speaker spatialized to left
and right sides at ±30°; **b** voice cue: male and female speakers spatialized to right side at
30°; **c** both cues: male and female speakers spatialized to left and rights sides at ±30°
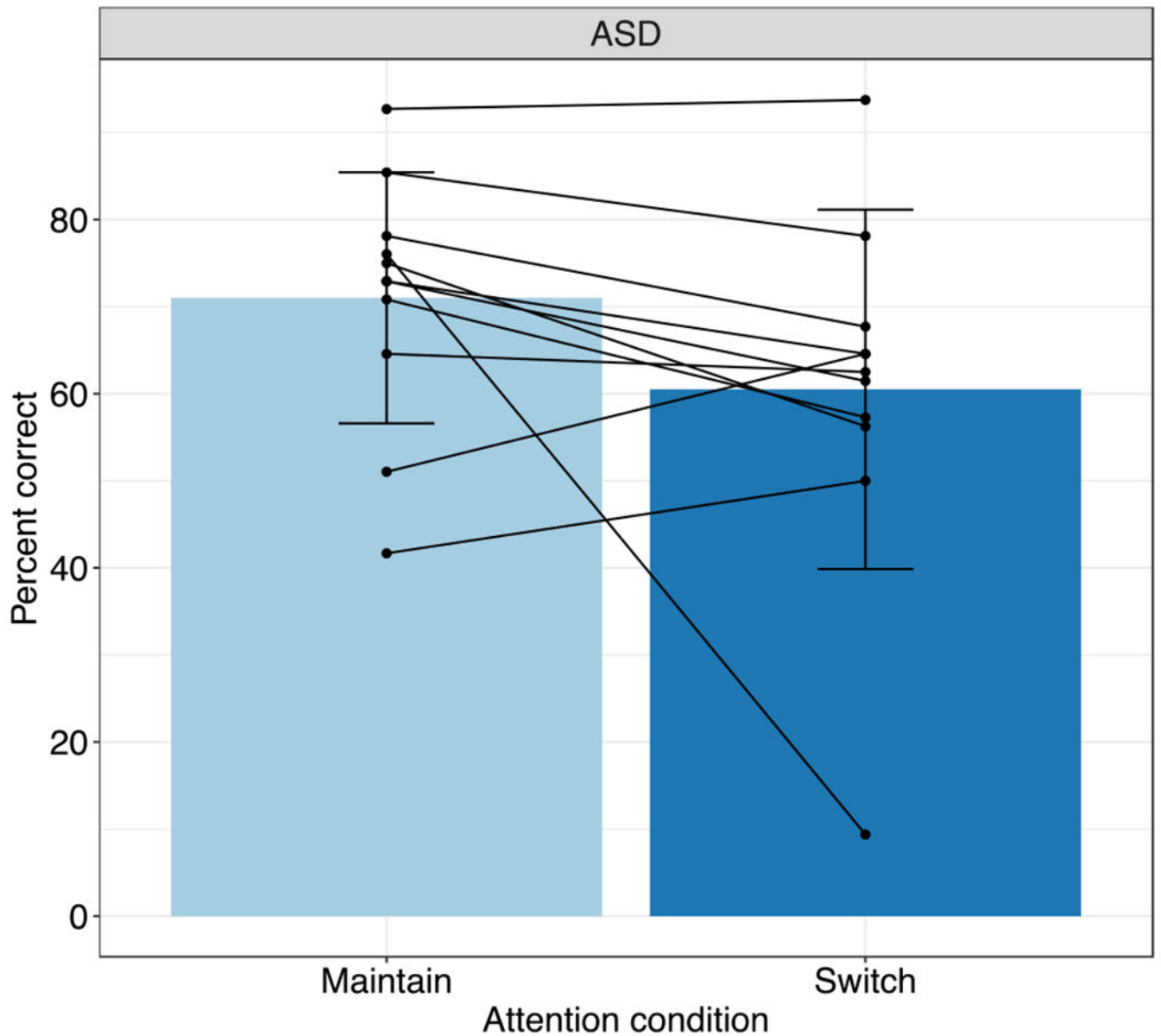
Fig. 3.
Task performance as a function of attention condition in the ASD group. Mean ± SD shown with bars; Individual data points shown with solid lines. No difference in performance on maintain-attention compared to switch-attention trials for the ASD group (light blue vs. dark blue bars)
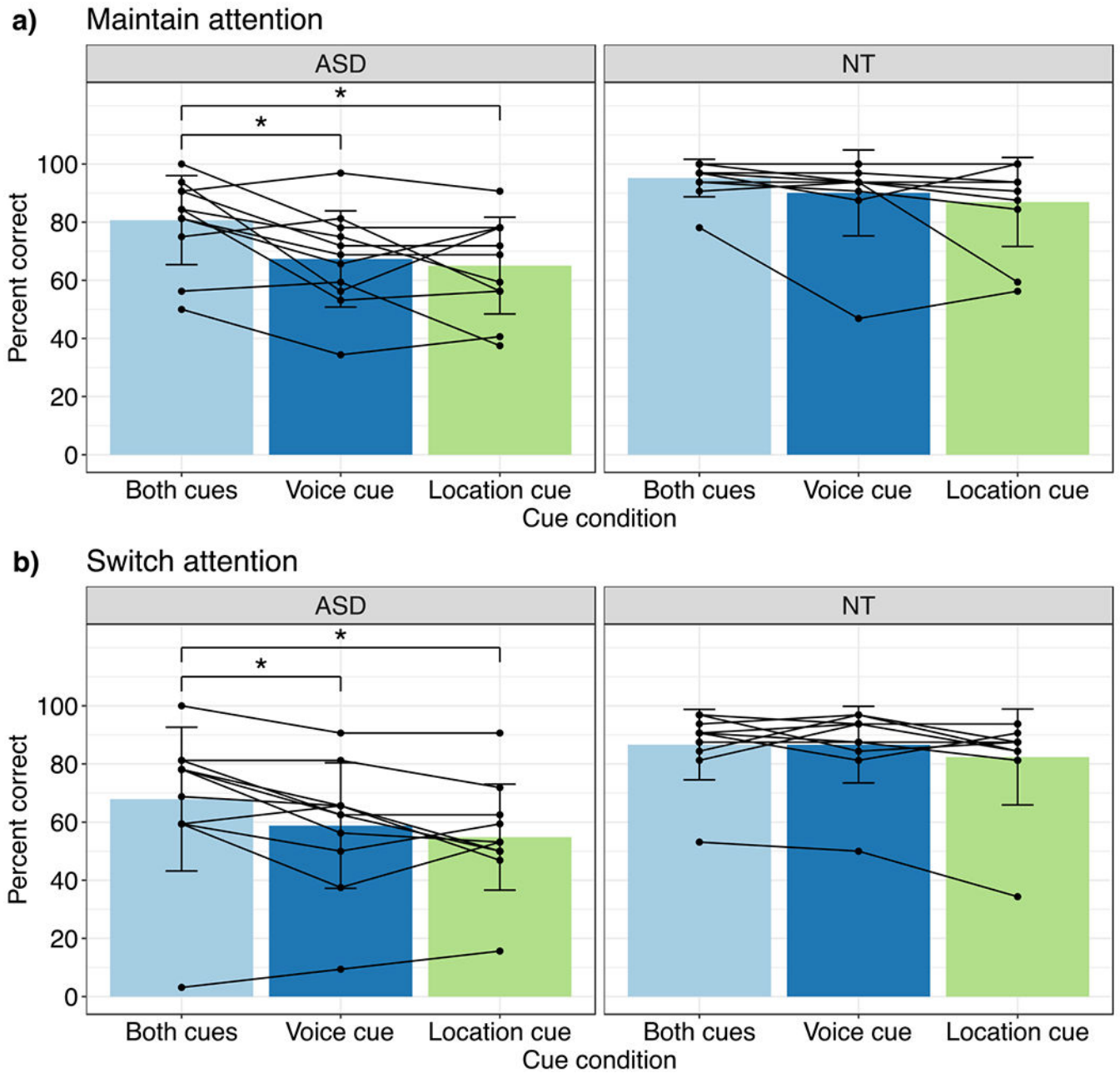
Fig. 4.
a Performance on Maintain-Attention trials as a function of Cue Condition in ASD and NT groups. Mean ± SD shown with bars; Individual data points shown with solid lines. * indicates p   .017 (Bonferroni adjusted a). Better performance on trials with both cues (left: light blue bar) compared to one cue for the ASD group. No difference in performance on trials with both cues compared to one cue for the NT group. **b** Performance on Switch-Attention trials as a function of Cue Condition in ASD and NT groups. Better performance on trials with both cues (left: light blue bars) compared to one cue for the ASD group. No

difference in performance on trials with both cues compared to just one cue for the NT group

Table 1

Sample characteristics

| | Group means | | | | Comparison statistics | |
| | ASD (*n* = 11; 4 female) | | NT (*n* = 11; 4 female) | | | |
	Mean (SD)	Range	Mean (SD)	Range	t	p
Age in years	21.75 (0.64)	21.00–23.29	21.79 (0.54)	21.07–22.75	− 0.14	0.893
Male	21.86 (0.75)	21.00–23.29	22.01 (0.53)	21.30–22.75		
Female	21.57 (0.42)	21.09–22.03	21.40 (0.33)	21.07–21.73		
WASI-II FSIQ-4	102.36 (14.33)	76–127	118.00 (6.63)	109–131	− 3.29	0.004
Male	102.29 (18.18)	76–127	116.43 (5.94)	109–123		
Female	102.50 (4.80)	96–107	120.75 (7.76)	113–131		

Comparison statistics are given by 2-tailed independent t-test

ASD autism spectrum disorder; *WASI-II FSIQ-4* Wechsler Abbreviated Scales of Intelligence, Second Edition—Full Scale Intelligence Quotient-4 Subtests