Auditory stream segregation in children with Asperger syndrome

T. Lepistö a,b,*, A. Kuitunen a, E. Sussman c, S. Saalasti a,b, E. Jansson-Verkasalo d,e, T. Nieminen-von Wendt b,f, T. Kujala a

aCognitive Brain Research Unit, Department of Psychology, University of Helsinki, P.O. Box 9, Helsinki FIN-00014, Finland
bDepartment of Child Neurology, Helsinki University Central Hospital, Finland
cDepartment of Neuroscience, Albert Einstein College of Medicine, USA
dFaculty of Humanities, Logopedics, University of Oulu, Finland
eDepartment of Clinical Neurophysiology, Oulu University Hospital, Finland
fNeuropsychiatric Rehabilitation and Medical Centre NeuroMental, Helsinki, Finland

ARTICLE INFO

Article history:
Received 5 January 2009
Accepted 8 September 2009
Available online 12 September 2009

Keywords:
Auditory
Asperger syndrome
Event-related potentials
Mismatch negativity
Speech perception
Streaming

ABSTRACT

Individuals with Asperger syndrome (AS) often have difficulties in perceiving speech in noisy environments. The present study investigated whether this might be explained by deficient auditory stream segregation ability, that is, by a more basic difficulty in separating simultaneous sound sources from each other. To this end, auditory event-related brain potentials were recorded from a group of school-aged children with AS and a group of age-matched controls using a paradigm specifically developed for studying stream segregation. Differences in the amplitudes of ERP components were found between groups only in the stream segregation conditions and not for simple feature discrimination. The results indicated that children with AS have difficulties in segregating concurrent sound streams, which ultimately may contribute to the difficulties in speech-in-noise perception.

1. Introduction

Asperger syndrome (AS) is an autism spectrum disorder which is characterized by deficient social interaction and communication skills, stereotyped behaviors, as well as circumscribed special interests (American Psychiatric Association, 1994; World Health Organization, 1993). Although language delay is not a general characteristic of AS (Volkmar and Klin, 2000), individuals with AS may have initial mild deficits in their language development, and have semantic–pragmatic deficits in their language use (Cederlund and Gillberg, 2004; Gillberg and Coleman, 2000; Wing, 1981). Furthermore, auditory hypersensitivity is common in AS (Dunn et al., 2002; Gillberg and Coleman, 2000), and might be partly explained by the enhanced cortical sound discrimination ability reported in this syndrome (Kujala et al., 2007; Lepistö et al., 2006, 2007).

In clinical practice, difficulties in perceiving speech in noisy situations are often reported by individuals with AS or their parents. However, research on this topic has so far been quite sparse. Alcantara et al. (2004) showed that adults with autism spectrum disorders required a higher signal-to-noise ratio in order to be able to perceive and repeat sentences during various kinds of background noise. This result was later replicated, and extended to the repetition of nonsense syllables (Alcantara et al., 2006). Teder-Sälejärvi et al. (2005) used both behavioral methods and event-related brain potentials (ERP) to investigate whether high-functioning adults with autism are able to focus on sounds coming from a certain spatial location, while ignoring similar sounds coming from different locations. The subjects heard eight streams of complex sounds at the same time, each coming from different spatial locations, and were instructed to press a response button to target sounds coming from a designated location. The subjects with autism were found to be less able to focus on the relevant sound source, and it was therefore concluded that their auditory selective attention skills were deficient. However, because the ability to segregate the sounds to multiple streams was inherent in this task, an alternative interpretation of this result is that there was an underlying impairment in auditory streaming abilities.

The fundamental ability to distinguish between different speakers and to hear a single voice in a crowd is referred to as auditory scene analysis (Bregman, 1990). Auditory scene analysis involves the ability to segregate sounds that come from different sources, and integrate those that belong together. The stream segregation process relies on a variety of acoustic cues, such as sound frequency and stimulation rate. In general, sounds with
similar acoustic features are heard as coming from a single source, whereas sounds that are acoustically distant from each other are perceptually segregated and perceived as originating from different sources.

In the present study, auditory event-related potentials (ERP) were used to investigate whether school-age children with AS have deficits in their auditory sound segregation abilities as compared to typically developing children. Specifically, the mismatch negativity (MMN) component of the ERP was used as the index of stream segregation. For the MMN elicitation, no behavioral response is required, and MMN can be recorded even when the subject is ignoring the sound stimuli (Naätänen et al., 1993; Paavilainen et al., 1993). Therefore, using the MMN, it is possible to study auditory stream segregation without the interfering influence of attention-based abilities.

MMN reflects early cortical stages of sound discrimination, and is elicited by perceptibly different sounds ("deviants") embedded into a sequence of repetitive sounds ("standards") (Naätänen et al., 1978; for a review, see Naätänen et al., 2007). In other words, the MMN is elicited when an incoming sound does not match with a sensory memory trace formed by the standard sound (Naätänen, 1992). Importantly, MMN elicited under passive conditions is associated with behavioral discrimination abilities, as its amplitude and latency to a particular sound contrast closely parallel the individual's discrimination ability of that contrast (Amenedo and Escura, 2000; Kujala et al., 2001; Lang et al., 1990; Novitski et al., 2004).

The MMN is often recorded to occasional sound changes in a sequence of repeated standard sounds (Oddball condition; Fig. 1A). When stream segregation is studied, intervening sounds are embedded in the original Oddball sequence (e.g., Sussman, 2005; Sussman et al., 2001; Winkler et al., 2003). The paradigm is set up so that the segregation of sounds to streams is a prerequisite for the detection of within-stream deviants. In other words, when the intervening sounds are far in frequency from the Oddball sequence, they segregate from each other, and the MMN is elicited by the deviant sounds in the Oddball sequence (Fig. 1C). In contrast, when the intervening sounds are near in frequency, they integrate with the Oddball sequence, and no MMN is elicited (Fig. 1B).

Importantly, both in children and adults the MMN elicitation in this paradigm has been shown to correspond with the subject's perception of the sounds as one integrated or two segregated streams in a separate behavioral study (Sussman et al., 2001; Sussman and Steinschneider, 2009). Therefore, the elicitation of the MMN can be used as a marker for determining whether stream segregation occurred.

In the present study, MMN was expected to be elicited by deviant tones in both the Oddball and Segregated conditions but not in the Integrated condition, as suggested by previous studies using the same paradigm (Sussman et al., 2001). Children with AS were expected to have diminished or absent MMN responses in the Segregated condition. This result would be consistent with less efficient stream segregation abilities.

2. Methods

2.1. Participants

The participants were 16 children with AS (mean age 8.10 years, range 7.0–9.10, 13 boys) and 14 typically developing children (mean age 8.10 years, range 7.6–9.7, 12 boys). The children with AS had been diagnosed by experienced clinicians according to the ICD-10 (World Health Organization, 1993) and DSM-IV (American Psychiatric Association, 1994) criteria at the Helsinki University Central Hospital or at the Helsinki Asperger Center (Dextra Medical Center). Their mean VIQ was 113 (range 90–145), and mean PIQ 106 (range 80–134) as assessed with WISC-III (Wechsler, 1991). They had no co-morbid diagnoses, and were all unmedicated at the time of testing.

A group of children without AS served as controls. They had no past or present neurological disorders, language or learning difficulties, and no emotional problems. Additionally, none of the children had close relatives diagnosed with autism spectrum disorder or other neurodevelopmental or psychiatric disorders. The mean VIQ of the control group on the WISC-III was 112 (range 96–137), and mean PIQ 109 (range 85–136).

The two groups did not differ in age (t(28) = 0.06, n.s.), VIQ (t(28) = 0.27, n.s.), or PIQ (t(28) = 0.47, n.s.). One child with AS was left-handed, and one control child was ambidextrous; all other children were right-handed.

The study was conducted in accordance with the Declaration of Helsinki, and was approved by the Ethical Committees of the Department of Psychology, University of Helsinki, and the Helsinki University Central Hospital. Informed written consent was obtained from the parents and assent from the children.

2.2. Stimuli and paradigm

The experiment consisted of three conditions (Oddball, Segregated, Integrated; Fig. 1). In each condition, the stimuli were pure tones of 50 ms (including 5 ms rise/fall times).

In the Oddball condition, there was one repetitive standard sound (440 Hz, 52 dB SPL) that was randomly replaced with a deviant sound of higher intensity (67 dB SPL, p = .12; Fig. 1A). There was always at least one standard sound in between the deviants. The stimulus onset asynchrony (SOA) was 300 ms, and the sound sequence consisted of altogether 2000 stimuli.

The Segregated condition was otherwise similar to the Oddball sequence, except that two tones were presented between the consecutive tones of the Oddball condition, resulting in an SOA of 100 ms (Fig. 1C). These intervening tones were 2637 Hz tones, occurring with a random equiprobable distribution of four intensity values (47, 57, 62, and 72 dB SPL). Note that the intensity values of the intervening tones spanned above and below the intensities of the Oddball sequence. Thus, the intensity of the Oddball sequence had neither the loudest nor the softest sounds in the sequence. The resulting two simultaneous sound streams shared the same intensity range but were separable on the basis of their frequency distance (440 Hz vs. 2637 Hz). Segregation of the sounds by frequency allows the Oddball intensity (67 dB) to be detected as a deviant. The Segregated condition consisted of 6000 stimuli.

The Integrated condition was similar to the Segregated condition, except that the frequency value of the intervening tones (523 Hz) was near to that of the Oddball sequence (Fig. 1B); thus, the sounds were integrated into a single sound stream. The Integrated condition consisted of 6000 stimuli. It should be noted that although the SOA was shorter in the Segregated and Integrated conditions than in the Oddball condition, “Oddball sequence” tones occurred once every 300 ms in every experimental condition (Fig. 1).

Fig. 1. A schematic illustration showing segments of the three experimental conditions. Rectangles represent tones; their y-axis coordinate shows the tone frequency. Different intensity values are marked with different shades of gray. Note that the pace at which the “Oddball sequence” tones occur is once every 300 ms in each experimental condition.
The present paradigm makes it possible to obtain an objective physiological index of stream segregation. When the sounds are integrated into a single stream, the intensity variation of the intervening tones obscures the regularity of the standard intensity tone in the Oddball stream from being detected, which prevents MMN from being elicited (Rahne et al., 2007; Sussman et al., 2001; Sussman and Steinieschneider, 2006). In contrast, when the sounds segregate by frequency, the Oddball tones emerge as a separate stream, allowing the detection of the louder intensity Oddball tones and the elicitation of the MMN. Thus, when MMN is elicited by the deviant sounds in the Oddball sequence, we can conclude that these tones were physiologically segregated from the intervening tones.

Control conditions to delineate the MMN. For each experimental condition, a corresponding control condition was conducted that was identical to its respective experimental condition, except for the standard and the deviant sounds being exchanged so that the standard was 67 dB SPL, and the deviant 52 dB SPL. This way the ERPs to the 67 dB SPL tones when they were deviants in each experimental condition could be compared with the ERPs to the 67 dB SPL tones when they were standards in the corresponding control condition. The purpose of this was to control for the different stimulus effects on the obligatory responses elicited by the standard and deviant stimuli that may occur when presenting a louder sound. The Oddball control condition included 500 stimuli, and the other control conditions included 1500 stimuli.

The experiment was carried out in an electrically shielded and sound-attenuated chamber. During the experiment, the children were instructed to watch a self-selected soundless video and ignore the sound stimuli presented through four loudspeakers, two of which were located on the right side, and two on the left side of the TV screen where attention was directed. The children were video-monitored throughout the experiment, and accompanied by a parent if necessary. The order of presentation of the conditions was counterbalanced across participants.

2.3. ERP recording

High-density EEG (amplified by BioSemi ActiveTwo amplifiers, band pass DC-67 Hz, sampling rate 256 Hz) was recorded using a 64-location electrode cap and additional electrodes at mastoids. Eye movements were monitored with electrodes placed below and at the outer corner of both eyes. EEG-epochs of 700 ms (including a 100-ms pre-stimulus time) were offline averaged separately for each stimulus type and condition. Epochs contaminated by artifacts causing peak-to-peak deflections exceeding 100 μV in any channel were excluded. The epochs were digitally filtered with 1–20 Hz band-pass filter and baseline-corrected with respect to the 100-ms pre-stimulus period. The data were re-referenced to the average of the left and right mastoid recordings. The final data set consisted of, on the average, 197 (range 139–238) accepted deviant trials per condition for controls, and 195 (range 113–234) for children with Asperger syndrome.

2.4. Data analysis

First, given that MMN is evoked by a deviant sound, it was important to determine that there were no significant differences in the ERP responses evoked by the standard stimuli (e.g., the P1 and N2 components in children). The peaks of the standard-ERP responses obtained from the control conditions were measured. The P1 peak latency was identified at 50–150 ms, and the N2 peak latency at 200–400 ms from stimulus onset. The P1 and N2 amplitudes (integrated over 50 ms) were measured at the latencies of their individual peak amplitudes at the Fcz electrode. The significance of each peak was assessed at the Fcz electrode with a two-tailed t-test against zero. Group differences were studied by one-way ANOVAs for repeated measures.

The MMN is delineated by subtracting the ERP elicited by the standard from the ERP elicited by the deviant. These deviant-minus-standard-ERP difference waveforms were constructed for each condition using the deviant-ERP obtained from the experimental condition and the standard-ERP obtained from the control condition. The MMN peak latency was identified at 100–300 ms from stimulus onset from these waveforms. The MMN amplitudes (integrated over 50 ms) were measured at the latencies of their individual peak amplitudes at the Fcz electrode. The MMN mean amplitudes were also measured at the grand-average peak latencies at the Fcz electrode; these data were used for the testing of the statistical presence of the MMN in different conditions before group comparisons.

The statistical presence of the MMN was determined in each condition at the Fcz electrode on the difference waveforms using a two-tailed t-test to verify whether the mean amplitude was significantly different from zero. Mixed model analysis of variance (ANOVA) for repeated measures was used to compare amplitude across conditions and groups.
groups. The between-subjects factor was group (AS vs. Control) and the within-subjects factor was condition (Oddball vs. Segregated). An additional analysis of scalp distribution of the MMN was conducted using a mixed model ANOVA for repeated measures (Group \times Condition \times Electrode) including the electrodes F3, F4, Fc3, Fc4, C3, C4, Cp3, Cp4, P3, and P4. The MMN latencies were analyzed at Fcz with two-way (Group \times Condition) ANOVA. The Greenhouse–Geisser correction was applied when appropriate. Newman–Keuls post hoc test was performed to calculate the sources of significant main effects and interactions.

Fig. 4. Grand averaged ERP waveforms elicited by control tones (thick line) and deviant tones (thin line) in the Segregated condition, shown for all the 64 electrodes used.
3. Results

The control tones elicited P1 and N2 responses in the Oddball and Segregated conditions (Fig. 2). However, in the Integrated condition only P1 was distinguishable. The N2 was suppressed due to the fast stimulus rate, and thus was not analyzed (Fig. 2; cf. Sussman et al., 2008). A significant P1 component was elicited in both groups in the Oddball condition (controls: t(13) = -11.28, p < .0001; children with AS: t(15) = -13.00, p < .0001) and in the Segregated condition (controls: t(13) = -7.34, p < .0001; children with AS: t(15) = -5.78, p < .0001). In the Integrated condition a significant P1 was elicited only in the control group (controls: t(13) = -2.37, p < .04; children with AS: t(15) = -1.54, n.s.). A significant N2 component was elicited in both groups in the Oddball condition (controls: t(13) = 5.88, p < .0001; children with AS: t(15) = 3.51, p < .01) and in the Segregated condition (controls: t(13) = 6.37, p < .0001; children with AS: t(15) = 7.71, p < .0001). No significant group differences were found for the P1 amplitude (Oddball condition: F(1,28) < 1, n.s.; Segregated condition: F(1,28) < 1, n.s.; Integrated condition: F(1,28) < 1, n.s.) or for the N2 amplitude (Oddball condition: F(1,28) = 3.55, n.s.; Segregated condition: F(1,28) < 1, n.s.).

A statistically significant MMN was elicited in both groups of children by deviant sounds in the Oddball condition (controls: t(13) = 5.23, p < .001; children with AS: t(15) = 5.23, p < .001), and in the Segregated condition (controls: t(13) = 5.96, p < .002; children with AS: t(15) = 5.43, p < .001), but not in the Integrated condition (controls: (t(14) = 0.99, n.s., children with AS (t(16) = 0.58, n.s.), which is consistent with previous child studies (e.g., Sussman et al., 2001) (Table 1, Figs. 3–5). As the MMN was not significantly elicited in the Integrated condition for either group of children, this condition was not further analyzed for comparisons of MMN amplitude and latency.

The main result was that the MMN amplitude in the Segregated condition was significantly smaller in the group of children with AS compared with the control group (interaction between group and condition: F(1,28) = 6.11, p < .02) (Table 1, Figs. 3–5). In contrast, the MMN amplitude did not significantly differ between groups in the Oddball Condition. Further, there was no significant group difference in the MMN scalp distribution (Fig. 6) or latency (Table 1).

4. Discussion

In the present study, the ability of children with AS to segregate sound streams based on frequency separation was investigated using auditory event-related brain potentials. Previous studies have shown that the elicitation of the MMN can be used as an index of stream segregation, since being able to detect changes in the Oddball sequence is dependent on successful streaming (e.g., Sussman, 2005; Sussman and Steinschneider, 2006). The main finding of the present study was a diminished MMN amplitude elicited by deviant
The study was supported by the Finnish Cultural Foundation, the Finnish Academy of Science and Letters, the Swedish Research Council, the European Union (FP5), and the National Institutes of Health. The work was also supported by the University of Helsinki, the Academy of Finland (grant number 128840), and the National Institutes of Health (DC 06003).

**References**


Lepisto¨, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., Na¨a¨ta¨nen, R., 2005, 2006, 2007. Therefore, it is interesting to note that children with AS in the current study had difficulties in using frequency proximity as a cue for streaming. However, it has been shown that the frequency discrimination acuity has no direct correlation with the stream segregation ability (Rose and Moore, 2005; Sussman et al., 2007). For example, although 9–11 years old children are as able as adults to discriminate frequency differences, their stream segregation abilities are not yet adult-like (Sussman et al., 2007). This is probably because in addition to the precise coding of acoustic features, stream segregation involves more general perceptual and cognitive processes, such as attention, integration of prior context, and schematic knowledge (Snyder and Alain, 2007). That is, stream segregation is a higher-level complex skill perhaps affected more by experience than by the ability to discriminate sound frequencies (Sussman et al., 2007).

This preliminary study investigating complex auditory skills in children with AS used non-speech stimuli, and a paradigm in which the sounds would be clearly perceived as either one or two streams (Sussman et al., 2007). Therefore, caution should be taken in drawing conclusions, and clearly, further studies are needed. Nonetheless, the present results are consistent with the notion that deficient passive stream segregation abilities may in part contribute to difficulties in perceiving speech in noisy environments reported in individuals with AS. The finding has important implications, since speech perception rarely takes place in quiet conditions without any background noise.

**Acknowledgements**

We thank all the children and their families for participation. The study was supported by the Finnish Cultural Foundation, the University of Helsinki, the Academy of Finland (grant number 128840), and the National Institutes of Health (DC 06003).