**REPORT**

**Evidence for reduced domain-specificity in auditory processing in autism**

Anna Järvinen-Pasley and Pamela Heaton

*Goldsmiths College, University of London, UK*

**Abstract**

Neurological and behavioral findings indicate that atypical auditory processing characterizes autism. The present study tested the hypothesis that auditory processing is less domain-specific in autism than in typical development. Participants with autism and controls completed a pitch sequence discrimination task in which same/different judgments of music and/or speech stimulus pairs were made. A signal detection analysis showed no difference in pitch sensitivity across conditions in the autism group, while controls exhibited significantly poorer performance in conditions incorporating speech. The results are largely consistent with perceptual theories of autism, which propose that a processing bias towards featural/low-level information characterizes the disorder, as well as supporting the notion that such individuals exhibit selective attention to a limited number of simultaneously presented cues.

**Introduction**

Autism is a pervasive neurodevelopmental disorder affecting brain function. Although it is commonly associated with varying degrees of intellectual impairment, the characteristic neurocognitive profile is uneven, and ‘unimpaired’ and enhanced cognitive skills are frequently documented. The disorder is diagnosed on the basis of deficits in social interaction, communication, and cognitive flexibility, noted before the age of 3 years (American Psychiatric Association, 1994). Given that the diagnostic criteria are primarily social, it is unsurprising that ‘preserved’ skills are most commonly observed in domains that do not involve social interaction. Thus, while research in autism has highlighted difficulties in understanding pitch-mediated linguistic cues in speech, or prosody (Kujala, Lepistö, Nieminen-von Wendt, Näätänen & Näätänen, 2005; see McCann & Peppé, 2003, for a review), and deficits in semantic processing (Dunn, Vaughan, Kreuzer & Kurtzberg, 1999; Rapin & Dunn, 2003), relatively intact or superior musical pitch processing has also been widely reported (Applebaum, Egel, Koegel & Imhoff, 1979; Bonnel, Mottron, Peretz, Trudel, Gallun & Bonnel, 2003; Heaton, 2003, 2005; Heaton, Hermelin & Pring, 1998; Heaton, Pring & Hermelin, 1999; Mottron, Peretz & Ménard, 2000).

Neurological investigations have provided further evidence for an atypical auditory information-processing profile in autism. Consistent with the behavioral findings (e.g. Heaton et al., 1998; McCann & Peppé, 2003), this work has identified a pattern in which the neural processing of speech and vocal sounds, but not tonal and environmental sounds, appears abnormal in individuals with autism relative to typical controls (e.g. Ceponienė, Lepistö, Shestakova, Vanhala, Alku, Näätänen & Yaguchi, 2003; Gervais, Belin, Boddaert, Leboyer, Coez, Sfaiello, Barthélémy, Brunelle, Samson & Zilbovicius, 2004). These studies have prompted the suggestion that autism may be characterized by a speech-specific attentional deficit, and increased attention to non-speech information. This may then result in enhanced processing of pitch variations in speech and, correspondingly, diminished appreciation of linguistically functional pitch, i.e. prosody (Gervais et al., 2004).

Theories of autism that have attempted to address the question of co-occurring assets and deficits have tended to focus upon atypical perceptual processing. The earliest of these accounts is the weak central coherence (WCC)
theory (Frith, 1989; Happé, 1999; Happé & Frith, 2006), which predicts that in individuals with autism, the typically observed propensity to process perceptual, visuospatial-constructional, and verbal-semantic stimuli globally and in context is weakened, while their performance in tasks in which a featural/surface-biased information processing style conveys an advantage will be superior relative to individuals without the disorder (Happé, 1999). Alternatively, the theory of enhanced perceptual functioning (EPF; Mottron & Burack, 2001; Mottron, Dawson, Soulières, Hubert & Burack, 2006) proposes that neural networks underpinning perceptual processing are ‘overspecialized’ and predispose locally oriented and enhanced perceptual functioning in autism. Although perception may play an unusually dominant role in high-level cognitive tasks in autism, higher-order functions are assumed to be unimpaired.

The present study was designed to test the hypothesis that children with autism show similar sensitivity to pitch in both speech and non-speech stimuli. In order to test this hypothesis, auditory processing was compared across speech and music domains. Pitch sequence pairs that either shared exact four-pitch sequences or differed in two pitches only were presented for same/different discriminations. The experimental paradigm comprised three conditions of increasing complexity: those of music–music, speech–speech, and speech–music stimulus pairs, respectively. Because the task was to determine whether the melody was the same or different in each contrast, no between-group differences will be predicted in the music–music condition due to the stimuli varying on only one dimension (perceptual). By contrast, it is hypothesized that children with autism will show superior perceptual performance over their typical controls in the more complex conditions incorporating speech, as they include competing perceptual and linguistic information. Specifically, the speech–speech condition requires the participants to ignore semantic and other linguistically relevant perceptual elements to solve the task. The speech–music condition is more complex still by requiring participants to compare pitch sequences across auditory information belonging to different stimulus classes.

Method

Participants

Nineteen children with a formal diagnosis of autistic disorder or Asperger disorder according to DSM-IV criteria (APA, 1994) participated in this study. They were recruited from two specialist educational establishments for children with high-functioning autism and Asperger syndrome. The diagnostic information was gathered from school files of documented medical diagnoses and clinical reports. The selected children all met the following criteria: both their British Picture Vocabulary Scale (BPVS; Dunn, Whetten & Pintilie, 1997) and Raven’s Standard Progressive Matrices (RSPM; Raven, Court & Raven, 1992) standardized scores fell within the intellectually unimpaired range (> 70), they had a mono-lingual English-speaking home environment, and no evidence of a hearing or a neurological impairment. The children with autism were individually matched to typically developing children for chronological age (CA), as well as verbal and non-verbal intelligence. The children in the control group were recruited from a mainstream primary and a secondary school. These children all met the same criteria as specified for children with autism. Children were screened for musical training, and those who had undergone periods of extensive musical training (two or more years of individual music lessons) were excluded from the study. The rationale for the latter exclusion criterion was that as training increases auditory analytic abilities within the musical domain, this ensured that all participants had ‘typical auditory experience’.

Table 1 shows the demographic characteristics of the two groups of children. No significant between-group differences in age, the BPVS, or the RSPM standardized scores are in evidence (t-tests all $p \geq .463$).

Visual training stimuli

The aim of the training trials was to familiarize participants with the same/different discrimination paradigm. As the aim of the study was to identify group-specific biases in auditory processing, it was important that performance in the experimental trials would not be influenced by the content of the preceding training trials. Previous studies have successfully utilized visual training prior to auditory discrimination testing (e.g. Peppé & McCann, 2003) and such a method was adopted here. Ten pairs of visual figures were used to train participants. In half of the pairs the two figures were of the same color, and in the remaining half of the pairs the two figures were of different color. In some of the pairs the figures were identical although the colors may not have been, while in the remaining pairs the degree of difference between the two figures was variable. The purpose of using pairs in which the colors of the shapes were different was to draw the participants’ attention to the fact that the shapes can be the same although the colors are not; this is an analogue of the speech–music condition used in the current experiment, where the pitch information can be the same across different stimulus classes. Correspondingly, the same color figure pairs were intended to
be an analogue of the music–music and speech–speech stimulus pairs used in the test stimuli.

**Auditory stimuli**

A list of frequently spoken words was compiled using the MRC (Medical Research Council) Psycholinguistic Database, Version 2.0 (Wilson, 1988). The selected words were 34 neutral nouns and adjectives, each with four syllables and a familiarity rating that was greater than 550 (max. 657). Multiple exemplars of each of the words, that sounded relatively naturalistic, were then read and recorded by a native English-speaking female. The exemplars differed in pitch trajectories and all syllables were equally stressed. The exact individual pitches, their mean fundamental frequencies ($F_0$) in Hz, pitch curves, and pitch separations together with the exact timings of the individual elements in the samples were then traced using the Melodyne software package (Neubäcker & Gehle, 2003). Stimuli selected for use in the ‘different’ condition included two same and two different pitches, maintained the same sequences, and ordered violating pitches in the second and third positions in the four-pitch/syllable sequences. Melodyne allows modification of samples, and minimal timing and pitch adjustments were made to six of the stimuli. An example of the musical composition of a different pair (‘responsible 1’ and ‘responsible 2’) is illustrated in Figure 1.

Twenty-four four-tone musical forms, which shared the same pitch and timing properties as the speech samples, were created using a Casio tone 202 electronic keyboard (acoustic piano setting) as follows. The four-pitch sequences of each of the speech stimuli were played and recorded using the GoldWave® software package. The melodies were then imported to Melodyne, where their exact pitch and timing characteristics were analyzed.

Table 1  Characteristics of the two participant groups (SDs; range, in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>CA (years)</th>
<th>Sex</th>
<th>VIQ (BPVS)</th>
<th>NVIQ (RSPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism group ($n=19$)</td>
<td>11.55 (2.34; 7.75–16.92)</td>
<td>16M; 3F</td>
<td>92 (17.39; 70–135)</td>
<td>94 (10.81; 70–119)</td>
</tr>
<tr>
<td>Control group ($n=19$)</td>
<td>11.68 (2.21; 7.92–15.85)</td>
<td>13M; 6F</td>
<td>95 (16.03; 78–124)</td>
<td>91 (14.31; 72–121)</td>
</tr>
</tbody>
</table>

Table 2  Means for the pitch of the first, second, third, and fourth syllable/tone, and the average pitch difference for critical syllables/tones (SD), in the pitch sequences of the speech and music stimuli ($F_0$ in Hz)

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>1st syllable/tone $M$</th>
<th>SD</th>
<th>2nd syllable/tone $M$</th>
<th>SD</th>
<th>3rd syllable/tone $M$</th>
<th>SD</th>
<th>4th syllable/tone $M$</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>404.21</td>
<td>61.90</td>
<td>282.00</td>
<td>107.62</td>
<td>267.38</td>
<td>61.50</td>
<td>324.88</td>
<td>143.51</td>
</tr>
<tr>
<td>Music</td>
<td>422.50</td>
<td>56.03</td>
<td>276.96</td>
<td>103.15</td>
<td>249.42</td>
<td>59.66</td>
<td>332.46</td>
<td>139.02</td>
</tr>
</tbody>
</table>

![Figure 1](image.png)  Example of the melodic composition of a ‘different’ pair (‘responsible 1’, top, and ‘responsible 2’, bottom; please note that the graphic representation displays the melodic composition an octave higher than the auditory stimuli).
using GoldWave®, and inserting the same auditory information twice into a sound file. Each pair was separated by a one-second gap. For the ‘different’ pairs, the two different melodic forms of the same words were inserted into sound files, again separated by one-second gaps. This ‘different’ condition shared half of the stimuli used in the ‘same’ pairs. The pairs for the music–music condition were created as described earlier, using the corresponding musical samples. For the speech–music condition, the speech preceded the musical form in half of the stimuli, and vice versa in the other half. Twelve same and 12 different pairs were created as described above. Thus, for example, for a ‘different’ pair where the speech segment was ‘responsible 1’, the musical form would be that of ‘responsible 2’. The procedure described above resulted in a total of 72 stimulus pairs. The order of the stimulus items was randomized with respect to stimulus class and whether the pairs were the same or different.

Procedure

The children were tested individually in a quiet room in their own school. In the initial training phase, children were presented with visual slides of figure pairs of the same shape and color. The experimenter asked the child to say what s/he noticed about the figures (same/different color; same/different shape). Once the child gave a correct answer, the pair where the two figures were most different was shown. The child was again asked to comment on them. The child was then presented with the remaining three slides in which the pairs of figures became progressively more similar. The experimenter corrected the child’s discrimination where necessary. The same procedure was then carried out for the slides in which the shapes were of different color, to draw the child’s attention towards the ‘figural’ similarities. Participants took as many practice trials as needed until reaching an 80% success rate. Ninety-five percent of children achieved this during the first trial, and the remaining children reached this criterion on a second trial. Once the visual training was complete, the experimenter explained: ‘Now we are going to do the same with things that we listen to. You are going to hear some pairs of sounds, and sometimes they will only be a little different, sometimes they will be slightly more different, and sometimes they will be the same. So, we are going to listen to pairs of sounds, and you will need to decide whether they sound the same or different. Sometimes we are going to hear a voice speaking and a piano playing, but that doesn’t necessarily mean that they will be different. In fact, they can be the same, and so you will need to listen carefully whether their tune is the same or not. Let’s listen to a pair, and please tell me whether you think the pairs have the same or a different tune.’ The child was presented with an example of each type of a stimulus pair, six items in total. The test stimuli were then presented via headphones. The experimenter played the first item of the test stimuli, and recorded the child’s response. The test was divided into two blocks of 36 items, with a 3-minute pause in between. To avoid ‘fatigue effects’, the order of presentation of the two stimulus blocks was counterbalanced across participants. The control children received the test blocks in the same order as their counterparts with autism.

Results

Individual raw scores for the music–music, speech–speech, and speech–music stimulus pairs were converted to signal detection measures of perceptual sensitivity (d′), and the response bias or decision criterion (C) (Green & Swets, 1966). Values for d′ are presented in Figure 2.

The participants’ d′ values for all experimental conditions were compared against chance-level performance (0) by applying one-sample t-tests. This analysis showed that the performance of the autism group was significantly above chance with all stimuli (music–music (t = 4.88, p < .001; 67.1% correct mean response rate); speech–speech: (t = 8.54, p < .001; 65.6% correct mean response rate); and speech–music: (t = 7.08, p < .001; 64.7% correct mean response rate)). The control group performed significantly above chance in the music–music (t = 5.59, p < .001; 69.5% correct mean response rate) and speech–speech (t = 2.40, p = .027; 55.3% correct mean response rate) conditions. However, their performance was at chance in the speech–music condition (t = −1.59, ns; 48.9% correct mean response rate). The data within each stimulus
category were normally distributed (Levene’s $F$-statistic for the variables: music–music: $p = .87$; speech–speech: $p = .15$; and speech–music: $p = .06$). A three by two repeated measures analysis of variance was carried out on the data, with stimulus type (music–music, speech–speech, and speech–music) as the within-subjects variable, and diagnosis (autistic/control) as the between-subjects variable. The results showed a significant main effect of stimulus type ($F(2, 35) = 10.16, p < .001$); diagnosis ($F(1, 36) = 10.48, p = .003$); and a significant stimulus type by diagnosis interaction ($F(2, 35) = 7.58, p = .002$). When this analysis was performed with BPVS and RSPM standardized scores entered as co-variates, the results remained essentially unchanged (stimulus type: ($F(2, 33) = 4.07, p = .026$); diagnosis: ($F(1, 34) = 10.16, p = .003$); and stimulus type by diagnosis interaction: ($F(2, 33) = 8.82, p = .001$)).

Pair-wise comparisons of stimulus type with Bonferroni $\alpha$ adjustments were carried out for each group. For children with autism, there was no significant difference between any two levels of the repeated measures variable (music–music versus speech–speech: ($t(18) = .10, ns$); music–music versus speech–music: ($t(18) = .27, ns$); and speech–speech versus speech–music: ($t(18) = .51, ns$)). For controls, all comparisons were statistically significant (music–music versus speech–speech ($t(18) = -2.17, p = .044$); speech–speech versus speech–music ($t(18) = 2.92, p < .009$); music–music versus speech–music ($t(18) = 6.17, p < .001$)).

Between-group comparisons at each level of the repeated measures variable were calculated using Bonferroni $\alpha$ adjustments. This analysis showed no significant difference between groups for performance with the music–music stimuli ($t(36) = -.81, ns$). There was a significant between-groups difference in performance in the speech–speech condition ($t(36) = 2.09, p = .043$) and in the speech–music condition ($t(36) = 7.12, p < .001$).

Finally, in order to examine response biases, the $C$ values were plotted for each of the within-subjects variables. Figure 3 shows the mean $C$ statistic for both groups of participants across the three stimulus types.

Figure 3 shows that overall, children with autism showed weaker response biases as compared to the controls. While participants with autism were biased towards making ‘the same’ judgments in the music–music condition, they showed virtually unbiased performance in the crucial conditions that tested sensitivity to pitch in speech. The controls showed a bias in all conditions; in the within-domain speech–speech and music–music conditions they were biased towards making ‘different judgments’, whereas they showed a bias towards judging the cross-domain speech–music stimuli as ‘the same’.

**Discussion**

The current study tested the hypothesis that auditory processing in autism is characterized by reduced domain-specificity relative to that observed in CA and intelligence-matched typically developing controls. The results showed that children with autism were able to detect perceptual changes in speech stimuli as well as changes in musical stimuli. Although the controls showed similar levels of pitch sensitivity to the children with autism when...
comparing short musical pitch sequences, deterioration in performance was seen in the conditions that required the discrimination of speech. This deterioration was most apparent with across-domain stimuli that required pitch sequences to be compared across speech and music, thus providing support to our hypothesis. While the performance of the controls in the speech–speech condition was significantly above chance, reduced sensitivity to perceptual changes in these stimuli, together with those in the speech–music pairs, was found in both the between-group analyses and the response bias scores.

Two perceptual theories of autism were outlined in the introduction. The pattern of current results is consistent with the WCC theory (Frith, 1989; Happé, 1999; Happé & Frith, 2006), in that the enhanced perceptual processing of children with autism relative to controls was particularly evident with stimuli incorporating speech. The finding that both groups showed similar pitch sensitivity in the music–music condition suggests that perceptual processing abilities of the two groups did not differ generally, and that the observed differences reflected an autism-specific tendency to process speech at the perceptual level. This pattern of results is consistent with a weak drive for central coherence, although the current stimuli cannot elucidate global-level processes as conceptualized by the WCC theory. The results can also be interpreted within the context of the EPF theory (Mottron & Burack, 2001; Mottron et al., 2006), insofar as the participants with autism exhibited superior processing of low-level perceptual information in speech compared to controls. The results from the study are also consistent with the suggestion that cognitive processing in autism is not characterized by mandatory higher-order control, and that perceptual versus higher-order control can be regulated more flexibly than is the case in typical development. The current results failed to support the hypothesis of universally enhanced perceptual processing in autism, as the discrimination performance of the autism group did not significantly exceed that of controls in the music–music condition. This result is consistent with findings from two recent musical contour discrimination studies showing no processing advantage in autism relative to controls (Foxton, Stewart, Barnard, Rodgers, Young, O’Brien & Griffiths, 2003; Heaton, 2005).

Alternatively, the pattern of current results may reflect overly focused selective attention towards a limited number of available cues in complex stimuli (i.e. acoustic-perceptual over linguistically relevant information) in the autism group. Previous studies have reported an atypically narrow attentional focus for both complex speech (Schreibman, Kohlenberg & Britten, 1986) and visual (Koegel, Schreibman, Britten & Laatinen, 1979) stimuli in autism. In the study by Schreibman et al. (1986), either non-verbal or echolalic children with autism were trained on complex speech stimuli including both content and intonation components. The findings showed that, whereas the children with echolalia tended to respond selectively to the intonation component of the speech stimuli, the non-verbal children tended to selectively respond to the content component. By contrast, typical controls showed either selective responding to the content of the stimuli or no selective responding to either intonation or content components.

The performance of typically developing controls is consistent with these earlier findings in that both analysis and discrimination of small pitch differences in speech would clearly be hampered by overly selective attending towards the content. It may also be the case that attention to both content and intonation cues would limit the processing capacity/resources available for pitch analysis. Controls exhibited an opposite pattern of response biases in the within-domain speech–speech (and music–music) and across-domain speech–music conditions. However, their bias in the within-domain conditions was towards making ‘different’ judgments, and in the across-domain condition making ‘the same’ judgments, this suggests that the nature of the stimuli did not simply predispose them to judge pairs of words in the speech–speech condition (and melodies in the music–music condition) as ‘the same’, and pairs in the speech–music condition as ‘different’ without reference to their perceptual qualities. Instead, their performance was characterized by a high false alarm rate. This pattern of biases suggests that the controls did understand task demands and attempted to perform well (i.e. attend to the pitch information), but were ‘distracted’ by the linguistic relevance of the stimuli. The level of performance of the controls was significantly higher in within-domain conditions than in the more complex speech–music condition, where across-domain comparisons were required. Indeed, in this condition, performance scores were not significantly different from chance.

The finding that pitch information in speech has increased salience for children with autism is important given that deficits in prosodic and semantic processing are frequently described in high-functioning individuals with this disorder (e.g. Dunn et al., 1999; Kujala et al., 2005). It is also the case that between one-third and one-half of diagnosed individuals fail to acquire any expressive language during their lifetime (Bryson, 1996; Bailey, Phillips & Rutter, 1996; Lord & Paul, 1997). It may then be that overly selective attention towards auditory-perceptual features in speech hinders higher-level language processing and in some cases language acquisition and development in individuals with autism (cf. Schreibman et al., 1986). The finding that children with autism exhibited biased responding in the music–
music condition, and unbiased responding in the conditions incorporating speech, suggests that their perceptual discrimination ability was superior with speech as compared with music stimuli. Recently, Kuhl, Coffey-Corina, Padden and Dawson (2005) showed that mechanisms specialized for processing changes in vowel pitch had failed to develop in preschool children with autism who preferred a non-speech analogue to child-directed speech. In the current behavioral study, group differences in discrimination performance only emerged in conditions where specialized speech-processing mechanisms might disrupt the perception of changes in perceptual components in children without autism. It may thus be that auditory processing is more domain-general in autism than in typical development. The finding that the performance of the controls showed the most deterioration when across-domain (speech–music) stimuli were presented for discrimination is consistent with a domain-specificity account of auditory processing in typical development.

In conclusion, the results from this exploratory study showed that children with autism exhibited similar pitch sensitivity across different classes of auditory stimuli, resulting in superior awareness of speech pitch in comparison to typical controls. While overly focused selective attention towards auditory-perceptual cues may impact upon the processing of musical and other auditory information, its effects upon linguistic information processing may well be profound. Indeed, overly focused auditory processing in autism may contribute to the undercutting of emerging language skills. However, it may also be the case that enhanced awareness of the perceptual components in speech may aid, rather than hinder, language acquisition in autism, particularly in individuals with good non-verbal intellectual abilities. Future studies should further elucidate interpretations of the current data, for example, by utilizing stimuli that are free of semantic interference, such as those comprising nonsense syllables. Shorter syllable/tone sequences together with a shorter inter-stimulus interval would clarify the role of auditory working memory upon task performance. Studies should also relate perceptual discrimination abilities of individuals with autism to their higher-level language processing skills. Finally, a functional brain imaging study employing a similar paradigm to that used in the current study would enable the identification of neural processes supporting the auditory processing of pitch information present in speech and music in autism.

Acknowledgements

This work was submitted in partial fulfillment of a Doctor of Philosophy degree at the University of London by the first author. Dr Heaton’s work is supported by EU grant (12984) Stages in the Evolution and Development of Sign Use (SEDSU). We would like to thank Kristelle Hudry and Luke Smillie for their invaluable help with the data analysis, and three anonymous reviewers for their helpful comments on earlier versions of this manuscript. We are grateful to all the children who participated in this study, and their parents and teachers, for their kind cooperation.

References


Autism and reduced auditory domain-specificity


Received: 8 June 2006
Accepted: 4 December 2006