

# **Speech-in-noise recognition and the relation to vocal pitch perception in adults with autism spectrum disorder and typical development**

Stefanie Schelinski <sup>1,2</sup> & Katharina von Kriegstein <sup>1,2</sup>

<sup>1</sup> Technische Universität Dresden, Faculty of Psychology

<sup>2</sup> Max Planck Institute for Human Cognitive and Brain Sciences

## **Abstract**

We tested the ability to recognise speech-in-noise and its relation to the ability to discriminate vocal pitch in adults with high-functioning autism spectrum disorder (ASD) and typically developed adults (matched pairwise on age, sex, and IQ). Typically developed individuals understood speech in higher noise levels as compared to the ASD group. Within the control group but not within the ASD group, better speech-in-noise recognition abilities were significantly correlated with better vocal pitch discrimination abilities. Our results show that speech-in-noise recognition is restricted in people with ASD. We speculate that perceptual impairments such as difficulties in vocal pitch perception might be relevant in explaining these difficulties in ASD.

## **Keywords**

autism spectrum disorder, speech-in-noise, pitch, F0, auditory perception, speech perception

## Introduction

Recognising what another person is saying under noisy conditions, for example on a crowded street or when other persons are talking at the same time, is an everyday challenging experience. Noise significantly diminishes speech recognition and speech recognition becomes more difficult with increasing noise levels (e.g., Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumbly & Pollack, 1954).

Subjective reports (Alcantara, Weisblatt, Moore, & Bolton, 2004; Schelinski, Roswadowitz, & von Kriegstein, 2017) and behavioural study results (Alcantara et al., 2004; Foxe et al., 2015; Groen et al., 2009; Irwin, Tornatore, Brancazio, & Whalen, 2011; Smith & Bennetto, 2007) suggest that understanding speech in noisy environments (i.e., speech-in-noise perception) is altered or impaired in people with autism spectrum disorder (ASD). When presented with speech-in-noise, people with ASD had difficulties benefitting from temporal gaps in the competing speech or noise signal (such as brief pauses) (Alcantara et al., 2004; Groen et al., 2009). In typical developed adults as well as children, such temporal gaps in the competing speech or noise signal facilitate speech recognition (Alcantara et al., 2004; Duquesnoy, 1983a; Festen & Plomp, 1990; Groen et al., 2009). Surprisingly, in other speech-in-noise task conditions in which the speech signal was presented with continuous noise, the same ASD group individuals who did not benefit from temporary gaps in the competing speech or noise signal performed similar to the control groups (Alcantara et al., 2004; Groen et al., 2009). This result would suggest that it is not speech-in-noise recognition per se that is more difficult in ASD, but rather to take advantage of those parts of the speech stream that are not as noisy (i.e., to take advantage of temporary gaps in the noise). The first aim of the present paper was to re-evaluate speech-in-continuous noise perception in ASD with a paradigm that might be more sensitive to group differences than the ones used in the previous studies.

The second aim of the present paper was to test whether there is a relation between speech-in-noise recognition performance and the ability to perceive vocal pitch in typical development and in ASD. Speech-in-noise recognition relies on the perception and integration of a variety of acoustic voice features, such as the fundamental frequency (F0) (for reviews see e.g., Anderson & Kraus, 2010; Brown & Bacon, 2010). F0 is the lowest frequency within the speech signal and is perceived as vocal pitch (i.e., pitch is the

perceptual correlate of F0). Many previous studies investigated the impact of F0 cues on speech recognition when there are one or several competing speakers. These studies were conducted with people with and without hearing impairment (without reported autism) (Assmann & Summerfield, 1990; Bidelman & Yellamsetty, 2017; Brokx & Nootboom, 1982; Carroll, Tiaden, & Zeng, 2011; Drullman & Bronkhorst, 2004; Summers & Leek, 1998). In these studies, auditory speech streams (e.g., pairs of vowels) spoken by one or several speakers were presented at the same time. They showed that larger differences in F0 between these competing speech streams (i.e., how far apart the different speech streams were in frequency) facilitated speech recognition for the target speech. This suggested that F0 processing abilities might be important for understanding speech with competing speakers. Whether F0 processing abilities are also important for understanding speech-in-noise is largely unknown. One study, which tested adults with hearing impairment, found that the ability to discriminate non-vocal pitch (i.e., the F0 difference of two tones) correlated positively with speech-in-noise perception abilities (Glasberg & Moore, 1989). In ASD, non-vocal pitch perception abilities can be on a neurotypical level, while vocal pitch perception abilities are impaired (Jiang, Liu, Wan, & Jiang, 2015; Schelinski et al., 2017). Whether the vocal pitch perception impairment in ASD is associated with altered speech-in-noise perception is unknown.

To address our aims, we tested speech-in-noise recognition abilities when the speech signal was presented continuously with noise in adults with high-functioning ASD and pairwise matched typically developed control participants. We tested whether the ASD group would show lower speech-in-noise recognition performance as compared to typically developed individuals using a potentially more sensitive measure than has been used previously (Alcantara et al., 2004; Groen et al., 2009). For testing the relation between speech-in-noise recognition abilities and vocal pitch perception abilities, we used individual thresholds in a pitch discrimination test (Schelinski et al., 2017). Based on previous results (e.g., Brokx & Nootboom, 1982; Glasberg & Moore, 1989; Summers & Leek, 1998) we hypothesised a significant positive correlation between speech-in-noise recognition and vocal pitch discrimination within the control group, i.e., that better performance in speech-in-noise recognition would be associated with better performance in vocal pitch discrimination.

Investigating speech-in-noise recognition is important because auditory noise significantly impacts everyday functioning and can restrict communication (for review see Klatte, Bergstrom, & Lachmann, 2013; Picard & Bradley, 2001; Szalma & Hancock, 2011; van der Kruk et al., 2017). For example, limited recognition of what another person is saying can impede a fluent social conversation. This makes a conversation more challenging and also leads to enhanced experience of stress (see e.g., Mackersie & Cones, 2011; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017). An enhanced stress experience might in turn lead to avoidance of potentially noisy situations and places which is an often-reported phenomenon for people with ASD (e.g., Stiegler & Davis, 2010). Our study will contribute to understanding the underlying factors that explain difficulties in speech-in-noise perception in ASD. One such factor could be altered vocal pitch perception. Previous studies provide evidence that alterations in perceiving and interpreting basic sensory signals (e.g., voice acoustic characteristics) might be associated with difficulties in social cognition in ASD (e.g., Baum et al., 2015; Schelinski et al., 2017; Schelinski & von Kriegstein, 2019). By investigating the relation between vocal pitch perception and speech-in-noise perception, our study will contribute to a better understanding of these mechanisms. Thus, understanding speech-in-noise recognition mechanisms in typically developed individuals and ASD will likely enhance our knowledge about the underlying mechanisms of communication difficulties in ASD.

## **Methods**

### ***Participants***

We tested 16 adults with ASD (ASD group; Mean age = 33.75 years (range 20-51)) and 16 typically developed individuals (control group; Mean age = 33.69 years (range 18-52)). Groups were matched pairwise with respect to gender, chronological age, handedness, and full-scale intelligence quotient (IQ > 85; Mean Full-scale IQ: ASD group = 110.31 and control group = 111.50; Supplementary Table 1; for details see Schelinski et al., 2017 and Supplementary Methods). All participants had normal hearing abilities (hearing level equal or above 25 dB at the frequencies between 250 and 8,000 Hz; audiometer MADSEN Micromate 304, GN Otometrics, Denmark), no limitations or disorders associated with the ear

or hearing, and native German language. None of the participants took psychotropic medication. Participants in the ASD group had previously received a formal clinical diagnosis and underwent additional clinical assessment including the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000 and Autism Diagnostic Interview- Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994)). For a detailed participants description see Schelinski et al., 2017, Supplementary Methods and Supplementary Tables 1 and 2. All participants gave written informed consent in accordance with procedures approved by the Research Ethics Committee of the University of Leipzig.

### *Experiments*

The experiment included a test on speech-in-noise recognition and a test on vocal pitch discrimination. The tests were part of a larger study that also included tests on voice identity, vocal emotion recognition, non-speech vocal and non-vocal perception. The results of vocal pitch discrimination have been reported previously (Schelinski et al., 2017). Please see supplementary methods for information on experimental procedure for both tests.

#### *Speech-in-noise recognition test*

*Stimuli.* Stimuli consisted of 134 auditory-only two-word sentences spoken by one male speaker. The sentences were semantically neutral, phonologically and syntactically homogenous (the pronoun ‘er’ (‘he’) and a verb, e.g., ‘Er schreibt.’ (‘He writes.’)), and spoken in a neutral manner. During the experiment stimuli were mixed with speech-shaped noise (i.e., each stimulus was white noise filtered to have the same long-term average spectrum as the average of all the speech stimuli) created on-the-fly (i.e., speech-shaped-noise was created and mixed with the speech stimulus during the experiment, immediately before each trial, Kreitewolf, Mathias, & von Kriegstein, 2017). Speech-shaped noise has a comparable masking on the speech signal as adding a number of other speakers talking at the same time (‘multi-speaker-bubble’), simulating real-life noise situations (e.g., a noisy party; Brungart, Simpson, Ericson, & Scott, 2001). For a detailed stimulus description, see Supplementary Material.

*Experimental design.* We measured participants' individual speech reception thresholds (SRTs) using an adaptive tracking procedure (one up, one down staircase method; Kaernbach, 1991) such that the task difficulty of each trial was adapted to the responses in the preceding trial. In each trial, we presented one auditory-only sentence, which included a target word (e.g., 'writes'). Immediately after this auditory presentation, four written words (including the target word) were presented simultaneously on the computer screen. The alternative words were similar to the target word in their first phonemes. Participants had to decide which of the four words was presented in the preceding sentence (i.e., which word was the target word). Target words were presented in randomised order. The test contained five runs. Sentences were not repeated within a run. Responses were recorded using a response box. After each response, immediate feedback was provided. Here, the chosen word changed from white to green font if the response was correct or to red font if the response was incorrect. Before the test sessions, participants performed three practise trials. Completing the test took approximately 20 minutes.

*Estimation of SRTs.* The initial signal-to-noise ratio (SNR) was 10 dB. This value decreased in steps of 1.3 dB following a correct response and increased in steps of 4 dB following an incorrect response. After 4 reversals (defined as a switch from a correct to an incorrect response or vice versa within two consecutive trials), the up and down sizes were changed to a decrease in steps of 2 dB following a correct and an increase in steps of 0.7 dB following an incorrect response and the block of trials continued for further 10 reversals. We estimated the SNR in dB at which the participant responded with 75% accuracy. SRTs were estimated from a single run by taking the mean of all SNR values visited during the final 10 reversals. For estimating the individual SRTs, SRTs were averaged over the five runs.

### ***Vocal pitch discrimination test***

*Overview vocal pitch discrimination test:* To test the ability to recognise changes in vocal pitch, participants performed a vocal pitch discrimination test. Vocal pitch is the perceptual correlate of the vibration rate of the vocal folds of a speaker (Hanson & Chuang, 1999; Smith & Patterson, 2005; for an overview see Kreiman & Sidtis, 2011). Thus, in the vocal pitch discrimination test, the frequencies of the glottal pulse rates of the speech signal (fundamental frequencies; F0) were manipulated in order to simulate different pitches. F0 of a voice is the primary determinant of the perceived pitch. The mean F0 determines whether we perceive a voice as rather high or low. For example, the mean F0 for men is 115Hz and 220Hz for women (see Kreiman & Sidtis, 2011). That means that for an average female speaker, the vocal folds open and close 220 times per second and less for an average male speaker, i.e., 115 times per second. Therefore, we usually perceive an average female voice as higher in pitch than a male voice. Listeners can usually perceive very small differences of the mean F0 of as little as 2% (2.4 Hz, e.g., Smith, Patterson, Turner, Kawahara, & Irino, 2005). The stimuli consisted of five vowels (/a/, /e/, /i/, /o/, /u/) spoken by one male speaker (44100 Hz sampling rate, for a detailed description see Smith & Patterson, 2005; Smith et al., 2005). In each trial, we presented two stimuli (of the same vowel) successively and participants were instructed to decide which of the two stimuli had the higher pitch. We measured the individual thresholds where participants were still able to differentiate between two vocal speech stimuli based on the perceived F0 differences (just noticeable difference; JND) using an adaptive tracking procedure. One of the two stimuli always had a fixed F0 of 112 Hz (which is near to the average for man) and the other differed in F0 by an amount ( $\Delta F0$ ) defined in musical cents (1 semitone = 100 cents, whereby a semitone describes the difference in sound between two notes that are next to each other in the western musical scale).  $\Delta F0$  is a change in F0 from the first to second stimulus within a trial. For a detailed description of the vocal pitch discrimination test see Schelinski et al., 2017.

### *Statistical analyses*

We analysed the data using SPSS version 24 (IBM SPSS Statistics, NY, USA). Statistical tests were calculated two-tailed if not otherwise stated. We used one-tailed testing if we had an a priori directed hypothesis. The level of significance was defined at  $\alpha = .05$ . If not otherwise stated, all analyses included data from 16 participants with ASD and their pairwise matched controls. To test group differences we computed independent  $t$ -tests. For the speech-in-noise recognition test, we used the SRT (SNR in dB) and for the vocal pitch discrimination test, we used the threshold value for the JND in cent as dependent variables. We used Pearson's correlation coefficient for correlation analyses. We used Spearman's rho for not normally distributed variables (Shapiro-Wilk test). We identified outliers for each test and group separately. We defined outliers as cases who scored outside 1.5 times the interquartile range (Tukey, 1977), implemented as a standard procedure in SPSS version 24 (IBM SPSS Statistics, NY, USA). There were two outliers within the ASD group. One participant with ASD was an outlier because of low performance in the speech-in-noise recognition test and another participant with ASD because of low performance in the vocal pitch discrimination test. To test whether results from correlation analyses were driven by these outliers, we report correlation analyses including and excluding these two ASD participants. There were no further outliers in the ASD or the control group.

## **Results**

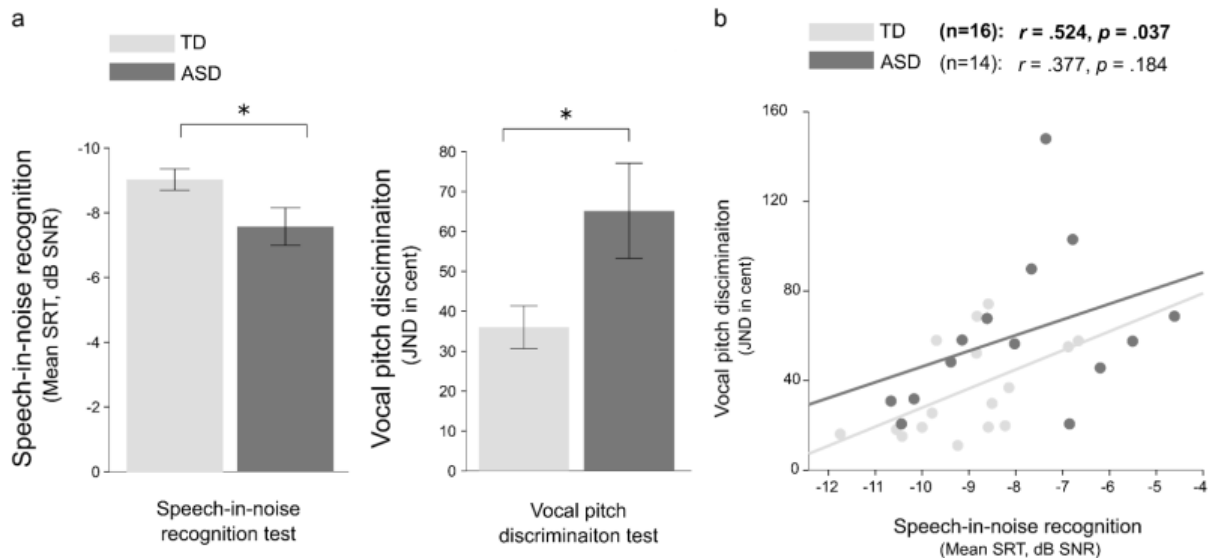
### *Significant group differences for speech-in-noise recognition*

The average SRTs were significantly different between the ASD and the control group ( $t(30) = -2.18$ ,  $p = 0.037$ ,  $d = 0.770$ ; Figure 1a). The ASD group had higher thresholds (-7.59 dB SNR;  $SD = +/- 2.32$ ;  $n = 16$ ) as compared to the control group (-9.04 dB SNR;  $SD = +/- 9.04$ ;  $n = 16$ ). Lower thresholds in the control group indicated better speech recognition performance in higher levels of noise. There were no significant group differences in the average number of trials presented throughout the test over the five runs (Mean control group = 253.69 trials; Mean ASD group = 249.38 trials;  $t(30) = 0.886$ ,  $p = 0.383$ ).



*Significant correlation between speech-in-noise recognition and vocal pitch discrimination within the control group*

The ASD group showed impaired vocal pitch discrimination abilities as compared to the control group (Schelinski et al., 2017; Supplementary Table 3). To test our hypothesis that better performance in speech-in-noise recognition would be associated with better performance in vocal pitch discrimination, we correlated individual performances in the speech-in-noise recognition test and the vocal pitch discrimination test within each group. Correlation analyses between the individual SRTs in the speech-in-noise recognition test and the JNDs in the vocal pitch discrimination test revealed a significant positive correlation within the control group ( $r = 0.582$ ,  $p = 0.009$ ,  $n = 16$ ; one-tailed; Spearman correlation; Figure 1b). This correlation indicated that within the control group, better vocal pitch discrimination abilities were associated with better performance in speech-in-noise recognition. Within the ASD group, there was no such significant correlation between the speech-in-noise recognition and the vocal pitch discrimination test ( $r = 0.451$ ,  $p = 0.106$ , for  $n = 14$ ;  $r = 0.135$ ,  $p = 0.617$ , for  $n = 16$ , Spearman correlation). The difference between the correlations between the two groups was not significant (Fisher's Z-test, one-tailed:  $Z = 0.438$ ,  $p = 0.331$  for  $n_{ASD} = 14$  and  $Z = 1.35$ ,  $p = 0.088$  for  $n_{ASD} = 16$ ).



*Figure 1*

Results of the speech-in-noise recognition and the vocal pitch discrimination tests (**a**) and the relation between speech-in-noise recognition and pitch discrimination abilities (**b**). **a** The ASD group performed worse than the typically developed control group (TD) in speech-in-noise recognition and vocal pitch discrimination. Note that lower SRTs indicate better speech recognition performance in higher noise levels (results for the vocal pitch discrimination test reported in Schelinski et al., 2017). **b** Within the control group, performance in the vocal pitch discrimination test correlated with performance accuracy in the speech-in-noise recognition test, implicating that better vocal pitch discrimination abilities were associated with better speech-in-noise recognition abilities. There was no such significant correlation within the ASD group. SRT = speech reception threshold; SNR = signal-to-noise ratio; JND = just noticeable difference. Error bars represent  $\pm 1$  SE; \*  $p < 0.05$ .

## Discussion

Our study on speech-in-noise recognition in high-functioning ASD revealed two key findings. First, the ASD group showed higher thresholds in speech-in-noise recognition than the control group. This means that typically developed individuals performed better in speech-in-noise recognition than individuals with ASD. Second, as hypothesised, speech-in-noise recognition abilities correlated with vocal pitch perception abilities in typically developed individuals, indicating that better vocal pitch perception abilities were associated with better speech-in-noise recognition abilities. There was no such correlation for individuals with high-functioning ASD, however the difference between the correlations in the two groups was not significant.

Our results show that people with ASD have difficulties in understanding speech in continuous noise. This contrasts previous studies where auditory-only speech-in-noise recognition difficulties were mainly restricted to speech-in-noise with temporal gaps in the noise signal (Alcantara et al., 2004; Groen et al., 2009; for studies on audio-visual speech-in-noise recognition see e.g., Foxe et al., 2015; Smith & Bennetto, 2007). Alcantara et al. (2004) found comparable SRTs between a group of adolescent and young adult people with ASD (mean age = 20.9 years +/- 11.1) when the speech stimuli (i.e., sentences including key words) were presented together with speech-shaped noise. Similarly, Groen et al. (2009) did not find significant group differences in SRTs between a group of children and adolescents with ASD (mean age = 14.0 years +/- 1.8) and age matched control group participants when the speech stimuli (i.e., single words) were presented together with pink noise. We assume that the difference between the previous and the present study results can most likely be explained by the threshold chosen for correct speech recognition performance. Alcantara et al. (2004) and Groen et al. (2009) used a threshold of 50% correct speech recognition performance to determine speech reception thresholds in an adaptive tracking design. That means that the thresholds refer to a performance where 50% of the speech could be correctly recognised. In contrast, in our study we used a higher recognition criterion of 75% correct responses. Other factors, such as differences in type of background noise, task difficulty or sample characteristics (e.g., age) could further contribute to the different study findings.

Individuals with ASD were able to understand speech on the same level as controls in a relatively high SNR (mean -7.59 dB SNR). This is in line with previous findings which indicate that speech recognition in ASD is unaffected when speech is presented in high SNRs (Schelinski, Riedel, & von Kriegstein, 2014 for SNR of 3 dB). This is also in line with our previous finding that the same ASD group as reported here performed on a neurotypical level in a speech recognition test in which the speech signal was presented in quiet, i.e., not additionally superimposed with noise (Schelinski, Borowiak, & von Kriegstein, 2016; also see e.g., de Gelder, Vroomen, & van der Heide, 1991; Woynaroski et al., 2013). These findings are in accordance with previous studies investigating another developmental condition (developmental dyslexia), which suggest that difficulties in speech-in-noise perception can be impaired without affecting speech recognition abilities in quiet (van Laarhoven, Keetels, Schakel, & Vroomen, 2018; Ziegler, Pech-Georgel, George, & Lorenzi, 2009). Note that group differences found in our study of 1.45 dB SNR are of significance as small differences in dB SNR indicate high functional difference in speech recognition. For example, for sentences presented in noise, a 1 dB change in SNR may correspond to as much as a 19% change in speech recognition for a steady background noise (Duquesnoy, 1983b; comp. Alcantara et al., 2004).

Previous studies indicated that processing F0 information is important for perceiving and recognising speech in challenging hearing conditions in typically developed individuals (e.g., Assmann & Summerfield, 1990; Bidelman & Yellamsetty, 2017; Brokx & Nootboom, 1982; Carroll et al., 2011; Drullman & Bronkhorst, 2004; Glasberg & Moore, 1989; Summers & Leek, 1998). In the majority of these studies, the importance of F0 in processing speech-in-noise was shown by manipulations of the speech material. For example, it was shown that listener's speech recognition performance for a target sentence benefits from increased differences in F0 between different competing speech streams (Assmann & Summerfield, 1990; Bidelman & Yellamsetty, 2017; Brokx & Nootboom, 1982; Carroll et al., 2011; Drullman & Bronkhorst, 2004; Summers & Leek, 1998). Results from one previous study indicate a direct relation between the ability to discriminate *non-vocal* pitch and speech-in-noise recognition performance (Glasberg & Moore, 1989). In the study by Glasberg and Moore (1989), discrimination abilities of the pitch

of complex tones were associated with speech recognition performance for speech embedded in speech-shaped noise in older adults with hearing impairment. We now provide evidence that also in typically developed adults with normal hearing abilities, the ability to recognise speech-in-noise is associated with the ability to perceive *vocal* pitch. Within the ASD group, vocal pitch perception abilities did not correlate significantly with speech-in-noise recognition performance. However, since there was no significant difference in correlation between the ASD and control group it is difficult to interpret this result at the present moment. If a replication study with a larger sample can refine differences in the correlation between vocal pitch perception abilities and speech-in-noise recognition performance between the two groups this would indicate a different processing mechanism for speech-in-noise perception in ASD in comparison to typically developed individuals.

There is evidence that executive functions are associated with speech-in-noise perception abilities in typically developed participant groups (for review see Dryden, Allen, Henshaw, & Heinrich, 2017) and that executive functions are dysfunctional in ASD (Demetriou et al., 2018). In our view dysfunctional executive functions are nevertheless not the most parsimonious explanation for our findings. First, the vocal pitch discrimination difficulties in the ASD group are difficult to explain by executive function differences between the groups, because in a previous study, the same ASD group who showed worse performance in the vocal pitch discrimination performed equally well as the control group in a test on vocal timbre discrimination. The vocal timbre discrimination test had exactly the same task design and task demands as the vocal pitch discrimination test (Schelinski et al., 2017). If the difficulties in vocal pitch discrimination in the ASD group were due to executive function differences between the groups, we would have expected similar impairments in the vocal timbre discrimination test. Thus, the vocal pitch recognition abilities are difficult to explain by potential executive function difficulties in the ASD in contrast to the control group. Second, another reason why we find it unlikely that executive function difficulties can explain the vocal pitch and speech-in-noise difficulties observed in the current study in ASD, is that in the ASD sample of our study there were no significant group differences between the control group and the ASD in tests on cognitive functioning that are related to executive functions, i.e., tests on working memory and attention

(Supplementary Table 1). A recent study showed that there are no specific subdomain functions of executive functions that are impaired in ASD (Demetriou et al., 2018), thus we expect that also other parts of executive function in our study sample are on a similar level as that of the controls.

A previous study showed that female talkers tend to be more intelligible than male talkers when speech is presented in different noise levels (Barker & Cooke, 2007). Here, we investigated speech-in-noise recognition abilities when participants listened to a male speaker. We used one speaker throughout the experiment, which is ecologically valid as listening to one speaker for an extended period of time is a common everyday listening situation. A systematic investigation of the influence of speaker gender and speaker variability on speech-in-noise perception abilities in people with ASD remains a subject to study.

Noise significantly impacts communication and difficulties in understanding speech-in-noise can impede communication and everyday functioning (for reviews see Klatte et al., 2013; Picard & Bradley, 2001; Szalma & Hancock, 2011; van der Kruk et al., 2017). For example, exposure to noise has negative effects on children's academic performance (Shield & Dockrell, 2008) and has been shown to significantly impact behaviour in children with ASD including speech recognition, listening behaviours, and listening stress (for reviews see Kanakri, Shepley, Varni, & Tassinari, 2017; van der Kruk et al., 2017). The speech-in-noise recognition difficulties found in the present study might explain some of the difficulties that people with ASD are faced with in auditory communication.

### **Acknowledgements**

We are grateful to our participants for taking part in the study. We thank Jens Kreitewolf for providing the speech-in-noise recognition test and Sam Matthias for providing the vocal pitch discrimination test.

### **Funding**

This work was funded by a Max Planck Research Group grant and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (SENSOCOM, grant agreement No 647051) to KvK.

## References

- Alcantara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107-1114.
- Anderson, S., & Kraus, N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *Journal of the American Academy of Audiology*, *21*(9), 575-585.
- Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels - vowels with different fundamental frequencies. *Journal of the Acoustical Society of America*, *88*(2), 680-697.
- Barker, J., & Cooke, M. (2007). Modelling speaker intelligibility in noise. *Speech Communication*, *49*(5), 402-417.
- Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology*, *134*, 140-160.
- Bidelman, G. M., & Yellamsetty, A. (2017). Noise and pitch interact during the cortical segregation of concurrent speech. *Hearing Research*, *351*, 34-44.
- Brox, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*(1), 23-36.
- Brown, C. A., & Bacon, S. P. (2010). Fundamental frequency and speech intelligibility in background noise. *Hearing Research*, *266*(1-2), 52-59.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *110*(5), 2527-2538.
- Carroll, J., Taden, S., & Zeng, F. G. (2011). Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. *Journal of the Acoustical Society of America*, *130*(4), 2054-2062.
- De Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69-86.
- Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., . . . Guastella, A. J. (2018). Autism spectrum disorders: A meta-analysis of executive function. *Molecular Psychiatry*, *23*(5), 1198-1204.
- Drullman, R., & Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *116*(5), 3090-3098.

- Duquesnoy, A. J. (1983a). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, 74(3), 739-743.
- Duquesnoy, A. J. (1983b). The intelligibility of sentences in quiet and in noise in aged listeners. *Journal of the Acoustical Society of America*, 74(4), 1136-1144.
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing*, 21, 1-27.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88(4), 1725-1736.
- Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., . . . Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, 25(2), 298-312.
- Glasberg, B. R., & Moore, B. C. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology*, 1-&.
- Groen, W. B., van Orsouw, L., ter Huurne, N., Swinkels, S., van der Gaag, R. J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, 39(5), 742-750.
- Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, 106(2), 1064-1077.
- Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face?. *Child Development*, 82(5), 1397-1403.
- Jiang, J., Liu, F., Wan, X., & Jiang, C. M. (2015). Perception of melodic contour and intonation in autism spectrum disorder: Evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders*, 45(7), 2067-2075.
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227-229.
- Kanakri, S. M., Shepley, M., Varni, J. W., & Tassinary, L. G. (2017). Noise and autism spectrum disorder in children: An exploratory survey. *Research in Developmental Disabilities*, 63, 85-94.
- Klatte, M., Bergstrom, K., & Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology*, 4, 578.
- Kreiman, J., Sidtis, D. (2011). *Foundations of Voice Studies*. Chichester, UK: Wiley-Blackwell.



- Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Frontiers in Psychology, 8*.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.
- 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.
- Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology, 22*(2), 113-122.
- McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology, 161*, 95-112.
- Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology, 40*(5), 221-244.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.
- Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience, 11*(11), 1812-1822.
- Schelinski, S., Riedel, P., & von Kriegstein, K. (2014). Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder. *Neuropsychologia, 65*, 1-11.
- Schelinski, S., Roswadowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research, 10*(1), 155-168.
- Schelinski, S. & von Kriegstein, K. (2019). The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders, 49*(1), 68-82.
- Shield, B. M., & Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *Journal of the Acoustical Society of America, 123*(1), 133-144.

- Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America*, *118*(5), 3177-3186.
- Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America*, *117*(1), 305-318.
- Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(8), 813-821.
- Stiegler, L. N., & Davis, R. (2010). Understanding sound sensitivity in individuals with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *25*(2), 67-75.
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212-215.
- Summers, V., & Leek, M. R. (1998). FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech Language and Hearing Research*, *41*(6), 1294-1306.
- Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682-707.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wisley Publishing Company.
- van der Kruk, Y., Wilson, W. J., Palghat, K., Downing, C., Harper-Hill, K., & Ashburner, J. (2017). Improved signal-to-noise ratio and classroom performance in children with autism spectrum disorder: a systematic review. *Review Journal of Autism and Developmental Disorders*, *4*(3), 243-253.
- van Laarhoven, T., Keetels, M., Schakel, L., & Vroomen, J. (2018). Audio-visual speech in noise perception in dyslexia. *Developmental Science*, *21*(1), e12504.
- Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(12), 2891-2902.
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732-745.