

## RESEARCH ARTICLE

# Altered processing of communication signals in the subcortical auditory sensory pathway in autism

Stefanie Schelinski<sup>1,2</sup>  | Alejandro Tabas<sup>1,2</sup>  | Katharina von Kriegstein<sup>1,2</sup> 

<sup>1</sup>Faculty of Psychology, Chair of Cognitive and Clinical Neuroscience, Technische Universität Dresden, Dresden, Germany

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

## Correspondence

Stefanie Schelinski, Faculty of Psychology, Chair of Cognitive and Clinical Neuroscience, Technische Universität Dresden, Bamberger Str. 7, 01187 Dresden, Germany.  
Email: stefanie.schelinski@tu-dresden.de

## Funding information

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 program (SENSOCOM, grant agreement No. 647051) to K.v.K. S.S. also received a research fellowship grant from the German Research Council (Deutsche Forschungsgemeinschaft [DFG], grant agreement No. 429525912).

## Abstract

Autism spectrum disorder (ASD) is characterised by social communication difficulties. These difficulties have been mainly explained by cognitive, motivational, and emotional alterations in ASD. The communication difficulties could, however, also be associated with altered sensory processing of communication signals. Here, we assessed the functional integrity of auditory sensory pathway nuclei in ASD in three independent functional magnetic resonance imaging experiments. We focused on two aspects of auditory communication that are impaired in ASD: voice identity perception, and recognising speech-in-noise. We found reduced processing in adults with ASD as compared to typically developed control groups (pairwise matched on sex, age, and full-scale IQ) in the central midbrain structure of the auditory pathway (inferior colliculus [IC]). The right IC responded less in the ASD as compared to the control group for voice identity, in contrast to speech recognition. The right IC also responded less in the ASD as compared to the control group when passively listening to vocal in contrast to non-vocal sounds. Within the control group, the left and right IC responded more when recognising speech-in-noise as compared to when recognising speech without additional noise. In the ASD group, this was only the case in the left, but not the right IC. The results show that communication signal processing in ASD is associated with reduced subcortical sensory functioning in the midbrain. The results highlight the importance of considering sensory processing alterations in explaining communication difficulties, which are at the core of ASD.

## KEYWORDS

auditory, autism spectrum disorder, inferior colliculus, sensory, speech-in-noise, vocal sounds, voice identity

## 1 | INTRODUCTION

Human communication requires the fast and accurate processing of sensory signals. For example, when we hear another person talking, the brain automatically extracts acoustic features of the voice and integrates them to successfully recognise what is said (speech recognition), who is speaking (voice identity recognition), and how the other

person feels (vocal emotion recognition) (Belin, Fecteau, & Bedard, 2004). Recently, it has been suggested that altered or impaired sensory processing might explain symptoms in clinical conditions with social communication difficulties, such as schizophrenia and autism spectrum disorder (ASD) (American Psychiatric Association [APA], 2013; Corlett et al., 2019; Gold et al., 2012; Javitt & Freedman, 2015; World Health Organization [WHO], 2004;

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

Schelinski, Roswadowitz, & von Kriegstein, 2017; Schelinski & von Kriegstein, 2019).

Traditionally, it is assumed that the differentiation of communication signals into speech, vocal identity, and emotional components occurs in the cerebral cortex and in limbic structures such as the amygdala (reviewed in Belin et al., 2004; Blank, Wieland, & von Kriegstein, 2014; Hickok & Poeppel, 2007; Schirmer & Kotz, 2006). Lesions in these structures can lead to relatively selective deficits in communication (Bonilha et al., 2017; Kummerer et al., 2013; Leff et al., 2009; Roswadowitz, Kappes, Obrig, & von Kriegstein, 2018; Scott et al., 1997; Sheppard et al., 2020; Van Lancker & Canter, 1982). Much less is known about the role of subcortical sensory pathway structures for communication impairments. Selective lesions in these pathways are extremely rare (Poliva et al., 2015) and tracking responses in tiny subcortical nuclei in humans in vivo is technically challenging (Forstmann, de Hollander, van Maanen, Alkemade, & Keuken, 2016). Studies on communication disorders therefore by-and-large ignore the potential role of subcortical sensory pathway structures for explaining communication difficulties (Bonilha et al., 2017; Kummerer et al., 2013; Leff et al., 2009; Roswadowitz, Schelinski, & von Kriegstein, 2017; but see Diaz, Hintz, Kiebel, & von Kriegstein, 2012; Gaebler et al., 2020).

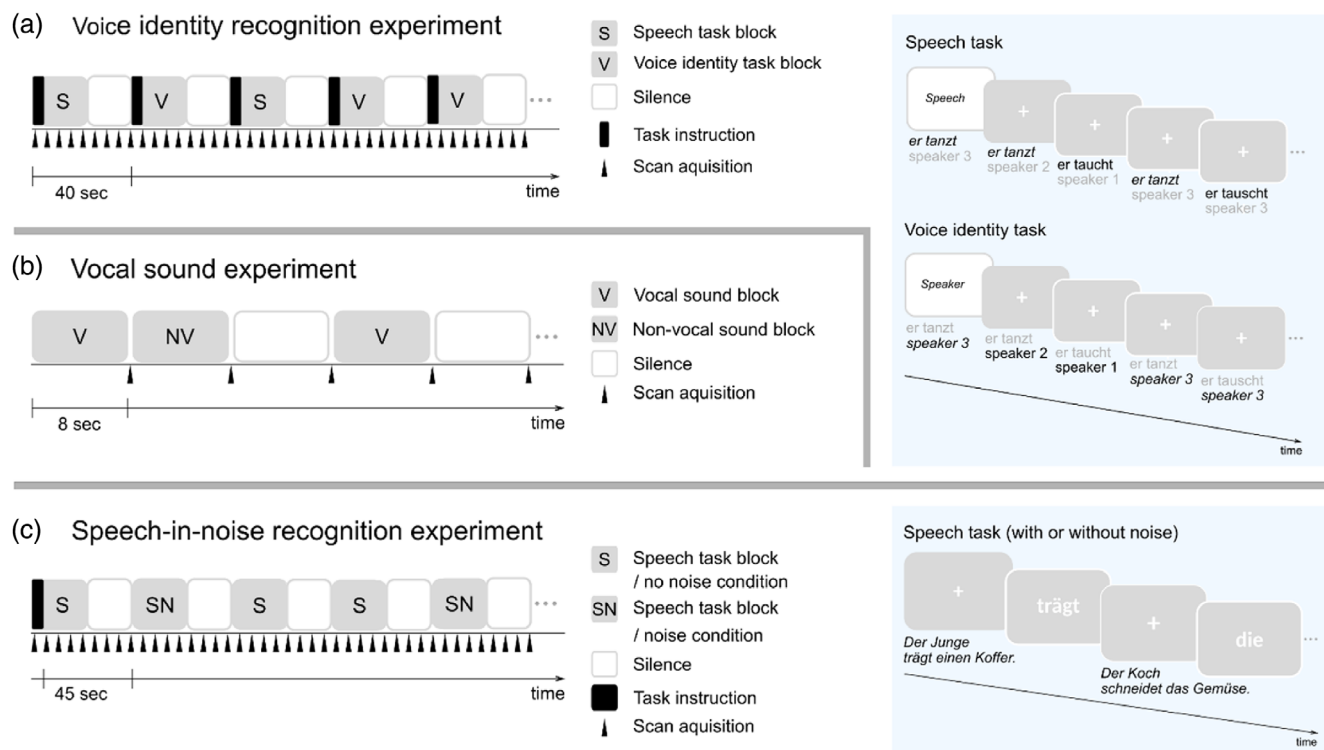
In ASD, it is to-date unclear at which stage communication difficulties arise. Many studies explain social communication difficulties, such as difficulties in recognising the identity of the voice, by alterations in social cognition such as social motivation rather than by altered sensory processing (Abrams et al., 2013; Abrams et al., 2019). In agreement with this suggestion, two previous studies found that communication difficulties in ASD are associated with intact auditory cortices, but dysfunctional connections to brain regions associated with reward and emotion processing (Abrams et al., 2013; Abrams et al., 2019). However, other studies in ASD show dysfunction in ASD of specific visual and auditory association cortices, such as middle temporal visual area 5 (V5/MT) (Borowiak, Schelinski, & von Kriegstein, 2018; Herrington et al., 2007), or dysfunction in the part of the temporal voice areas that process acoustic aspects of voices (Schelinski, Borowiak, & von Kriegstein, 2016). These findings support the notion that sensory processing could contribute significantly to difficulties in social communication in ASD (reviewed in Baum, Stevenson, & Wallace, 2015; Robertson & Baron-Cohen, 2017; Thye, Bednarz, Herringshaw, Sartin, & Kana, 2018). Whether such dysfunction arises in cerebral cortex *de-novo* or is also present in subcortical sensory nuclei is to-date unclear. A first indication that there might be subcortical sensory alterations in ASD comes from research on animal models (reviewed in Dadalco & Travers, 2018) and brainstem recordings (Russo et al., 2008; Russo, Nicol, Trommer, Zecker, & Kraus, 2009). Children with ASD as compared to typically developing children have altered brainstem responses for speech but not non-speech sounds (Russo et al., 2008; Russo et al., 2009). Whether these altered responses indicate altered processing in the subcortical auditory pathway is, however, unclear, as brainstem recordings suffer from low spatial resolution and potentially contributing cerebral cortex sources (Bidelman, 2018; Coffey et al., 2019).

The aim of the present study was to test whether, in ASD, voice identity and speech recognition in background noise are associated with altered processing in two nuclei of the auditory pathway that can be reliably assessed by functional magnetic resonance imaging (fMRI): the nuclei of the auditory midbrain (inferior colliculus [IC]) and thalamus (medial geniculate body [MGB]). We focused on voice-identity and speech recognition in background noise, because previous behavioural studies showed altered or impaired recognition abilities of voice identity (Boucher, Lewis, & Collis, 1998; Klin, 1991; Schelinski et al., 2017; Schelinski, Riedel, & von Kriegstein, 2014) and speech-in-noise (Alcantara, Weisblatt, Moore, & Bolton, 2004; Groen et al., 2009; Schelinski & von Kriegstein, 2020) in people with ASD in comparison to typically developed participants.

Groups of individuals with ASD and pair-wise matched controls participated in three fMRI experiments. First, participants performed tasks on speaker identity and speech recognition (voice identity recognition experiment; Figure 1a). Second, both groups passively listened to blocks of vocal and non-vocal sounds (vocal sound experiment; Figure 1b). In the third experiment, ASD and control participants performed speech recognition tasks on speech that was either presented with or without noise (speech-in-noise recognition experiment; Figure 1c). For the voice identity recognition and the speech-in-noise recognition experiment we recently showed dysfunctional processing of voice identity and speech-in-noise in the cerebral cortex (Schelinski et al., 2016; Schelinski & von Kriegstein, 2021), whereas processing in the cerebral cortex while passively listening to vocal sounds was on a neurotypical level (Schelinski et al., 2016).

We hypothesised that, if the ASD group has dysfunctional processing in subcortical sensory pathway nuclei compared to typically developed controls, this should be the case for tasks which are difficult for people with ASD and for which we found dysfunctional cerebral cortex processing in ASD (i.e., for voice identity vs. speech recognition in the voice identity recognition experiment (Schelinski et al., 2014; Schelinski et al., 2016; Schelinski et al., 2017), and speech-in-noise recognition versus speech recognition without noise in the speech-in-noise recognition experiment (Schelinski & von Kriegstein, 2021). Similarly, group differences in sensory subcortical nuclei function might also be present for passive listening of vocal versus non-vocal sounds (vocal sound experiment). This latter finding would speak for dysfunctional processing unrelated to a task.

Both, voice identity and speech-in-noise recognition abilities are key requirements for successful social communication (reviewed in, e.g., Blank et al., 2014; Mattys, Davis, Bradlow, & Scott, 2012; Scott, 2019). Familiarity with a voice leads to a familiarity feeling towards other people (Blank et al., 2014; Maguinness, Roswadowitz, & von Kriegstein, 2018) and familiar voices' speech is easier to recognise (Kreitewolf, Gaudrain, & von Kriegstein, 2014; Nygaard & Pisoni, 1998). Recognising speech in a noisy environment, such as in a classroom or in a busy canteen, is an everyday challenging task and noise exposure can impede communication (reviewed in Klatte, Bergstrom, & Lachmann, 2013; Picard & Bradley, 2001; Szalma & Hancock, 2011; van der Kruk et al., 2017). A better understanding of the dysfunctional brain mechanisms underlying these



**FIGURE 1** Experimental design and example trials (light blue background) of the three functional magnetic resonance imaging (fMRI) experiments. (a) In the voice identity recognition experiment, there were two task conditions (speech task, voice identity task) and stimuli consisted of blocks of auditory sentences. At the beginning of each block, a key-word (“Speaker” or “Speech”) on the screen instructed the participants to perform the voice identity or the speech task. In addition to the task instruction, a target sentence was presented. For the ensuing sentences, participants decided for each sentence whether it was spoken by the target speaker (voice identity task) or whether it matched the content of the target sentence (speech task). Stimuli in the voice identity and speech task blocks were the same. MR-scans were acquired continuously. (b) In the vocal sound experiment, participants listened to blocks of vocal sounds (V), non-vocal sounds (NV), and silence (white boxes). One brain volume was acquired after each block. There was no task besides listening to the stimuli. MR-scans were acquired after each block. (c) In the speech-in-noise recognition experiment, stimuli consisted of blocks of auditory sentences that were either presented with noise (noise condition) or without noise (no noise condition). Each condition was presented in separate blocks. Participants decided whether a written word presented on the screen appeared within the previously presented auditory sentence or not. MR-scans were acquired continuously

behavioural difficulties in ASD will contribute to a better understanding of communication difficulties that are characteristic for the condition.

## 2 | METHODS

The aim of the present study was to test whether voice identity, vocal sound, and speech-in-noise recognition are associated with altered processing in the IC and MGB.

### 2.1 | Participants

Sixteen adults with ASD (ASD group) and 16 typically developed adults (control group) participated in Experiments 1 and 2 (voice identity recognition and the vocal sound experiment; Table 1a). In Experiment 3 (speech-in-noise recognition experiment; Table 1b), 17 adults with ASD participated in the ASD group and 17 typically developed

adults participated in the control group. All participants were free of psychotropic medication. The groups in each experiment were matched pairwise: each control group participant was matched to one participant in the ASD group with respect to gender (male or female), chronological age (age difference within each participant pair  $\leq 3$  years), handedness (right or left as assessed by a standard questionnaire; Oldfield, 1971), and IQ (Table 1; full-scale IQ difference within each participant pair was maximally 1 *SD* [15 IQ points]). IQ was assessed using the German adapted version of the Wechsler Adult Intelligence Scale (Wechsler, 1997; German version by von Aster et al., 2006). All participants had an IQ within the normal range or above (IQ > 85), indicating that all participants were on a “high-functioning” cognitive level. Additionally, groups showed comparable concentration performances (d2 test of attention; Brickenkamp, 2002; Table 1). Experiments 1 and 2 included the same group samples. Experiment 3 included another sample for the ASD and the control groups. Eleven ASD and 6 control group individuals participated in all three experiments. Participants took part in Experiments 1 and 2 in one fMRI session at the same day as part of a comprehensive study

<b>(a) Voice identity recognition and vocal sound experiment</b>					
<b>Characteristic</b>	<b>ASD group (n = 16)</b>		<b>Control group (n = 16)</b>		
Gender	13 males, 3 females		13 males, 3 females		
Handedness <sup>a</sup>	14 right, 2 left		14 right, 2 left		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>
Age	33.75	10.12	33.69	9.58	.986
Range	20–51		18–52		
WAIS-III <sup>b</sup> scales					
Full-scale IQ	110.31	13.79	111.50	10.97	.789
Verbal IQ	110.75	12.35	108.75	12.59	.653
Performance IQ	107.38	17.55	112.69	9.59	.296
Working memory	108.63	2.22	108.00	3.76	.887
d2 test of attention <sup>c</sup>	104.19	8.61	106.06	3.41	.645
AQ <sup>d</sup>	39.81	6.61	14.13	4.77	<.001*
<b>(b) Speech-in-noise recognition experiment</b>					
<b>Characteristic</b>	<b>ASD group (n = 17)</b>		<b>Control group (n = 17)</b>		
Gender	14 males, 3 females		14 males, 3 females		
Handedness <sup>a</sup>	15 right, 2 left		15 right, 2 left		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>
Age	30.53	10.15	31.35	10.03	
Range	20–54		21–54		
WAIS-III <sup>b</sup> scales					
Full-scale IQ	110.65	11.68	114.18	12.55	.402
Verbal IQ	111.47	11.30	113.71	11.92	.579
Performance IQ	107.53	14.26	111.47	12.82	.403
Working memory	110.12	13.81	112.88	13.11	.554
d2 test of attention <sup>c</sup>	104.24	14.07	107.12	7.17	.457
AQ <sup>d</sup>	37.41	8.65	16.12	5.31	<.001*

Note: \*Significant group difference ( $p < .05$ ).

Abbreviations: *M*, mean; *p*, *p*-value; *SD*, standard deviation.

<sup>a</sup>Handedness was assessed using the Edinburgh handedness questionnaire (Oldfield, 1971).

<sup>b</sup>WAIS-III = Wechsler Adult Intelligence Scale, third version (Wechsler, 1997; German adapted version by von Aster, Neubauer, & Horn, 2006); *M* = 100; *SD* = 15.

<sup>c</sup>d2 test of attention (Brickenkamp, 2002); *M* = 100; *SD* = 10.

<sup>d</sup>AQ = autism spectrum quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; German version adapted from Freitag et al., 2007; [http://kriegstein.cbs.mpg.de/AQ/AQ\\_Deutsch\\_Schelinski.pdf](http://kriegstein.cbs.mpg.de/AQ/AQ_Deutsch_Schelinski.pdf)).

A total score of 32+ is considered a useful cut-off for distinguishing individuals who have clinically relevant levels of traits associated with autism spectrum (Baron-Cohen et al., 2001).

on voice perception (Schelinski et al., 2016; Schelinski et al., 2017). Experiment 3 was performed as part of another study which was performed approximately 2 years after the study on voice perception.

All participants were native German speakers. They reported normal hearing abilities and no limitations or disorders associated with the ear or hearing. Normal hearing abilities were confirmed with pure tone audiometry (hearing level equal or above 25 dB at the frequencies of 250; 500; 1,000; 1,500; 2,000; 3,000; 4,000; 6,000; and 8,000 Hz tested in each ear separately).

In both experiments, participants in the ASD group had previously received a formal clinical diagnosis of Asperger syndrome (11 males,

3 females in the voice identity and the vocal sound experiment, and 12 males, 3 females in the speech-in-noise recognition experiment) or childhood autism (2 males in both studies, verbal-IQ 100 and 119) according to the diagnostic criteria of the International and Statistical Classification of Diseases and Related Health Problems (ICD-10; WHO, 2004). Additionally, the diagnoses for all participants in the ASD group (except for one participant in Experiments 1 and 2, and another participant in Experiment 3) were corroborated with the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000; German version by Rühl, Bölte, Feineis-Matthews, & Poustka, 2004) and, if caregivers were available ( $n = 9$  in the voice identity

**TABLE 1** Characteristics of the ASD and the control group. (a) Voice identity recognition and vocal sound experiment (Schelinski et al., 2016). (b) Speech-in-noise recognition experiment (Schelinski & von Kriegstein, 2021). Each participant in the control group was matched with respect to chronological age, gender, IQ, and handedness to the profile of one ASD group participant

recognition experiment and  $n = 11$  in the speech-in-noise recognition experiment), additionally with the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994; German version by Bölte, Rühl, Schmötzer, & Poustka, 2003) and the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003; German version by Bölte & Poustka, 2006; Supplementary Table 1).

Participants in the control group reported to have no neurological or psychiatric history and no family history of ASD. None of the control participants exhibited a clinically relevant number of traits associated with ASD as assessed by the autism spectrum quotient (AQ; Baron-Cohen et al., 2001; German version adapted from Freitag et al., 2007; Table 1).

The studies were approved by the Ethics Committee of the Medical Faculty at the University Leipzig, Germany (299-12-14092012). All participants gave written informed consent in accordance with the Declaration of Helsinki and procedures approved by the Research Ethics Committee of the University of Leipzig. For details, see Supplementary Methods – Participants.

### 3 | EXPERIMENTS

For participants who never had MRI before, we conducted mock MRI to familiarise them with the MRI environment. We used Presentation software (Neurobehavioral Systems Inc.) to present stimuli and record responses. We presented stimuli during the fMRI experiments using an MR confon system (Mark II; MR confon, Germany). Design and raw data for the voice identity recognition and the vocal sound experiment are the same as described in Schelinski et al. (2016). The design from the voice identity and the vocal sound experiment was based on two standard approaches that are commonly used to identify brain responses to voice identity (Roswadowitz et al., 2014; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003) and to more general to vocal sound (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Gervais et al., 2004). Also see Supplementary Methods – Experimental procedure.

#### 3.1 | Voice identity recognition experiment

##### 3.1.1 | Stimuli

Stimuli consisted of auditory-only two-word sentences spoken by three professional male native German speakers (22, 25, and 26 years old) in a neutral manner. The sentences were semantically neutral, phonologically and syntactically homogeneous. They consisted of the pronoun “er” (“he”) and a verb, for example, “Er kauft.” (“He buys.”). All speakers were unfamiliar to the participants. For more details, see Supplementary Methods – Experiments.

##### 3.1.2 | Experimental design

The fMRI experiment included two conditions for which we presented exactly the same stimuli: a voice identity and a speech task

(Figure 1a). Each condition was presented in 18 blocks (36 blocks in total). At the beginning of each block, participants saw the word “Speech” or “Speaker” on the screen to inform them about which task to perform. At the same time, they heard a sentence spoken by one of the three speakers (target). This was followed by a stream of 12 two-word sentences (test sentences) spoken by one of the three speakers. The 12 sentences within one block consisted of three phonologically similar sentences (e.g., “Er sieht” (“He sees”), “Er siegt” (“He wins”), “Er singt” (“He sings”)) that were each repeated four times. The four repetitions of each test sentence were spoken by the three different speakers. In the voice identity task, participants memorised the target speaker and indicated for each sentence in the ensuing block whether it was spoken by the target speaker or not, independent of the content of the sentence. In the speech task, participants memorised the content of the target sentence and indicated in the ensuing block for each sentence whether it had the same content, independent of the voice identity. Each task included 216 trials (432 trials in total). Each trial was 1.5-s long. In each trial, one test sentence was presented for approximately 0.9 s, and the response window was open until the end of the trial. Between blocks, there was a silent period of 18 seconds in which a fixation cross was presented on the screen. Within the experiment, each block was presented twice: On one presentation, participants performed the voice identity task, and on the other the speech task. Blocks and trials within each block were presented in a randomised order. The number of target items varied between two and four across the blocks and was the same between conditions. All three speakers were presented the same amount of times as the target speaker (voice identity task) and spoke the same amount of target sentences (speech task). Responses were made via a button box using the target and middle finger of the dominant hand. The experiment took approximately 24 min.

##### 3.1.3 | Speaker and task familiarisation

Before fMRI-acquisition, participants were briefly familiarised with the three speakers and the task. For a detailed description see Supplementary Methods – Experiments. Stimuli used for familiarization were not used during the fMRI experiments.

#### 3.2 | Vocal sound experiment

##### 3.2.1 | Stimuli

Stimuli consisted of 60 blocks containing vocal sounds, non-vocal sounds or silence (20 blocks per condition; Belin et al., 2000). Sounds can be found in a public repository from the authors (<https://neuralbasesofcommunication.eu/download>; originally downloaded from <http://vnl.psy.gla.ac.uk/resources.php>). Each block was 8 s long. Blocks of vocal sounds included speech (e.g., words and foreign language) and non-speech sounds (e.g., laughs and sighs). Blocks of non-vocal stimuli included sounds from machines (e.g., car sounds), nature

(e.g., wind), animals (e.g., a birdsong) and musical instruments (e.g., a saxophone).

### 3.2.2 | Experimental design

During the experiment, blocks were presented in a randomised order (Figure 1b). Between each block, there was a pause of 2,850 ms for image acquisition. Participants were instructed to close their eyes and listen attentively. The fMRI-acquisition took approximately 12 min. After the experiment, participants wrote down, as accurately as possible, the names of all sounds they remembered hearing. Participants were not informed on this memory task before the data acquisition.

## 3.3 | Speech-in-noise recognition experiment

### 3.3.1 | Stimuli

Stimuli consisted of auditory-only five-word sentences spoken by six male native German speakers (25–31 years old) in a neutral manner. The sentences were semantically neutral, phonologically and syntactically homogeneous; for example: “Der Junge trägt einen Koffer” (“The boy carries a suitcase”), or “Der Koch schneidet das Gemüse” (“The cook cuts the vegetables”). The final stimulus set included 90 sentences for each speaker: 45 stimuli without noise and 45 stimuli combined with noise (signal-to-noise ratio [SNR] of –8 dB; linear 10 ms fade-in and fade-out). All speakers were unfamiliar to all participants. For more details see Supplementary Methods – Experiments.

### 3.3.2 | Experimental design

Before the fMRI experiment, participants were familiarised with the speakers and the task. For a detailed description see Supplementary Methods – Experiments. During the fMRI, participants performed speech recognition tasks on speech that was presented with additional noise (noise condition) or without (no noise condition) (Figure 1c). Each condition was presented in 18 blocks (36 blocks in total) of 9 trials. Within each block, only one condition was presented. In each trial, one sentence was spoken by one of the six speakers (324 trials in total). At the end of each sentence, a written word (target word) appeared on the screen and participants decided whether this word appeared within the previously heard sentence or not. The task was the same for the sentences with and without additional noise. The written word was presented for 1 second, immediately followed by the next trial. To avoid training effects for the sentences, different sentences were presented in the noise and the no noise condition. Whether a sentence was presented with or without noise was counterbalanced across participants. All sentences were repeated maximum two times within one condition. If sentences were repeated

within one condition, they were spoken by a different speaker. Between blocks there was a silent period of 18 seconds in which we presented a fixation cross on the screen. Including the silence period, the duration of one block was approximately 45 seconds. Sentences within each block were presented randomly. The order of blocks was presented randomly for each participant, but kept the same for each matched pair of ASD and control participants. Responses were made via a button box using the index and the middle finger of the dominant hand. Total fMRI-acquisition time was approximately 27 minutes.

## 3.4 | Image acquisition

MR-images were acquired on a 3 T Siemens Magnetom Verio (Siemens, Germany). We used the 12-channel head coil for acquisition of the functional data in order to fit the headphones used for stimulus presentation, and a 32-channel head coil for the acquisition of the structural data.

### 3.4.1 | Functional MRI

In the voice identity and the speech-in-noise recognition experiments, volumes were acquired continuously (TR = 2.81 s; voice identity recognition experiment: 507 volumes; speech-in-noise recognition experiment: 581 volumes). In the vocal sound experiment, one volume was acquired at the end of each block (TR = 11 s, 60 volumes in total) allowing stimulus presentation without MRI gradient noise (Hall et al., 1999; Whitehead & Armony, 2018). In all three experiments, we used a gradient-echo EPI (echo planar imaging) pulse sequence (TE = 30 ms; flip angle = 90°; FoV = 192 mm × 192 mm; 2 mm slice thickness; interslice gap = 1 mm resulting in a resolution of 3 mm isotropic; 42 axial slices; acquisition bandwidth = 1,954 Hz; whole brain coverage; ascending acquisition). For B0 field mapping, a pair of 2D gradient echo images (TR = 0.488 s, flip angle 60°, pixel bandwidth = 327 Hz/pixel, AC-PC oriented acquisition) with different echo times (TE1/TE2 = 4.92 ms/7.38 ms) was obtained (Jezzard & Balaban, 1995). These images were measured at the same slice locations and with the same voxel resolution and image size as the EPIs.

### 3.4.2 | Structural MRI

For anatomical images, we used a standard T1-weighted 3D magnetization-prepared rapid gradient echo sequence (Mugler & Brookeman, 1990). For details, see Supplementary Methods – Image acquisition.

## 3.5 | Data analyses

All analyses included data from 16 participants with ASD in the voice identity recognition and the vocal sound experiment and

17 participants with ASD in the speech-in-noise recognition experiment, and their respective matched control group participants.

### 3.5.1 | Behavioural data

For analysing behavioural data, we used SPSS software (version 24, IBM SPSS Statistics, New York, NY). We used R (R Core Team, 2021) for creating figures. For group comparisons, we used analyses of variance (ANOVAs) and independent *t*-tests. We used paired samples *t*-tests for within-group comparisons. All statistical tests were two-tailed. The level of significance was defined at  $\alpha = .05$ .

### 3.5.2 | MRI data

We analysed MRI data using standard procedures in SPM software (version 12, Wellcome Trust Centre for Neuroimaging, UCL, London, UK) in a Matlab environment (version 9.3, The MathWorks, Inc., Natick, MA). For pre-processing, images were realigned and unwarped. Anatomical images were coregistered to the mean of the functional images. Images were normalised to the Montreal Neurological Institute (MNI) standard stereotactic space and spatially smoothed with a Gaussian kernel of 4 mm full width at half maximum. For all analyses, statistical parametric maps were generated by modelling the evoked haemodynamic response for the different conditions as boxcar functions convolved with a synthetic haemodynamic response function using the general linear model (high-pass filter 128 s) (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007). For all experiments, we performed one-sample *t*-tests across the single-subject contrast images for within group analyses. For between group analyses, we used two-sample *t*-tests comparing the means of the single-subject contrast images from both groups.

#### *Design matrix voice identity recognition experiment*

We modelled the conditions “voice identity task,” “speech task,” and “instruction” at the first level. The contrasts of interest were “voice identity task > speech task” in each group separately and the interaction between task and group. To account for group and task differences in performance, we included the performance difference (percent correct) between the speech and voice identity task as a covariate of no interest for within and between group analyses.

#### *Design matrix vocal sound experiment*

We modelled the conditions “vocal sounds” and “non-vocal sounds.” The contrasts of interest were “vocal sounds > non-vocal sounds” in each group separately and the interaction between condition and group.

#### *Design matrix speech-in-noise recognition experiment*

We modelled the conditions “speech task noise” and “speech task no noise” at the first level. The contrasts of interest were “speech task noise > speech task no noise” in each group separately and the interaction between task and group. To account for differences in task difficulty in the speech-in-noise recognition experiment, we included the individual differences in task performance (percent correct) between the noise and the no noise condition as covariate of no interest for within and between group task comparisons.

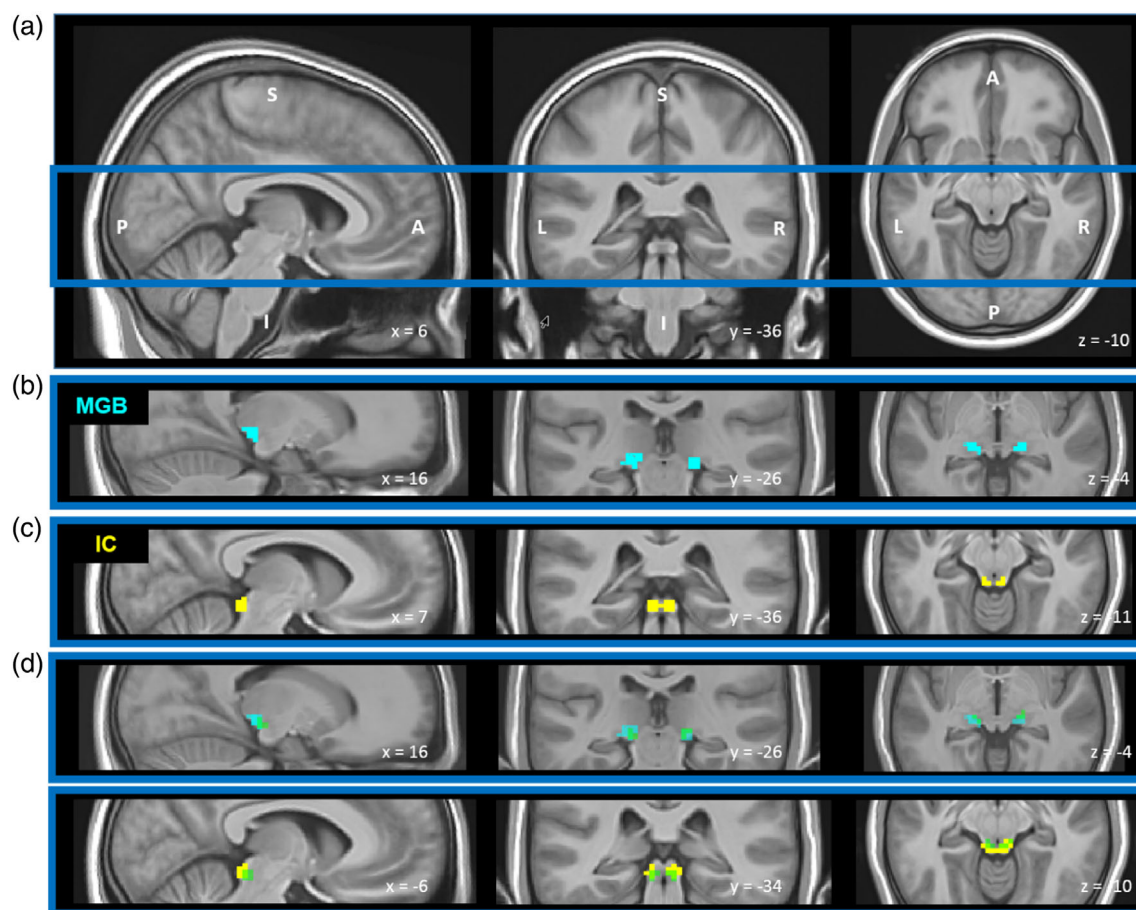
## 3.6 | Regions of interest

### 3.6.1 | Voice identity and vocal sound experiments

No independent functional localiser for the ASD and control group was available for these two experiments. Therefore, as a first approach we used ROIs provided in an independent atlas of the human subcortical auditory system that is based on functional MRI data (Sitek et al., 2019) (Figure 2a–c). In addition, to show robustness of the fMRI effects to different ROI definitions, we created ROIs defined as spheres centred on coordinates for the IC reported in previous functional studies (Gaebler et al., 2020; Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001; Supplementary Methods – Data analyses; Supplementary Figure 1).

### 3.6.2 | Speech-in-noise recognition experiment

For ROI definition in the speech-in-noise recognition experiment, we used data from participants who participated in both, the voice identity recognition and the speech-in-noise recognition experiment (11 ASD and 6 control group participants; Figure 2d). We used this approach to create functionally defined ROIs from the right and left IC and the right and left MGB. To create these ROIs, we used the contrast images from the overall response to speech stimuli (i.e., “voice identity task + speech task” contrasted against the implicit baseline) from the voice identity recognition experiment. We masked this contrast *t*-map with coarse anatomical regions for each of the ICs and MGBs (Supplementary Figure 2; Tabas, Mihai, Kiebel, Trampel, & von Kriegstein, 2020). We computed the IC and MGB ROIs by thresholding the masked *t*-maps at increasing thresholds until the ROI had a comparable volume of the IC and MGB reported in functional studies: 162 mm<sup>3</sup> for the IC (e.g., Amaral et al., 2016; Kang et al., 2008; Sitek et al., 2019) and 130 mm<sup>3</sup> for the MGB (e.g., Moerel, De Martino, Ugurbil, Yacoub, & Formisano, 2015; Figure 2d). The masking and thresholding was performed using FSLmaths (version 5.09). To test the validity of the functionally defined ROIs, we additionally report results using the functionally defined IC and MGB ROIs provided in an independent



**FIGURE 2** Overview of medial geniculate body (MGB) and inferior colliculus (IC) masks used as ROIs. Masks are plotted on a group mean structural image (experiment 1/2;  $n = 32$ ) (a). (b,c) For the voice identity recognition and the vocal sound experiment, we used masks from the MGB (b, cyan) and IC (c, yellow) provided in an independent atlas on the human subcortical auditory system (Sitek et al., 2019). (d) For the speech-in-noise recognition experiment, participant functionally defined MGB and IC masks were available. These were based on data from the voice identity recognition experiment (green) from individuals ( $n = 17$ ) who participated in both experiments (i.e., the speech-in-noise recognition and the voice identity recognition experiment). P, posterior; S, superior; A, anterior; I, inferior; L, left; R, right; x, y, z, coordinates in MNI space; MGB, medial geniculate body; IC, inferior colliculus

atlas of the human subcortical auditory system (Sitek et al., 2019) as used in the voice identity recognition and the vocal sound experiment (Supplementary Table 2).

### 3.7 | Significance threshold for fMRI data

We considered effects as significant at  $p < .05$  family wise error (FWE) corrected for the number of voxels in the ROI at peak level as implemented in SPM. The results were additionally Bonferroni corrected for four ROIs: the left and right MGB and left and right IC. Thus, the significance threshold was  $p < .0125$  FWE corrected for each ROI. Since Bonferroni correction has the risk of being overly conservative (White, van der Ende, & Nichols, 2019), we also report results that do not survive this correction. We indicate such results explicitly and consider them as less reliable. For information purposes only, all clusters at a whole brain level and a threshold of  $p < .001$

uncorrected are reported for the voice identity and the vocal speech experiment in Schelinski et al. (2016) and will be reported elsewhere for the speech-in-noise recognition experiment (Schelinski & von Kriegstein, 2021).

### 3.8 | Control analyses

We performed control analyses for potential group differences in head motion, task difficulty, and generalised blood oxygenation level-dependent (BOLD) responses (Supplementary Methods – Data analyses). We verified that our experimental setup had enough power to detect changes in the subcortical nuclei of interest by measuring the temporal SNR (tSNR) of the BOLD responses and controlled for potential group differences in tSNR (Supplementary Methods – Data analyses; Supplementary Results; Supplementary Figure 3; Supplementary Figure 4; Supplementary Table 3).



## 4 | RESULTS

### 4.1 | Voice identity and vocal sound experiment

#### 4.1.1 | Behavioural results

Behavioural results were reported and discussed previously (Schelinski et al., 2016). We here describe only the behavioural results that are relevant in the context of the present paper.

##### *Voice identity recognition experiment*

Performance in all tasks was >80% correct (Figure 3a; Supplementary Table 4). A repeated-measures ANOVA with the between-subject factor “group” (control, ASD) and the within-subject factor “task” (voice identity, speech) revealed a significant interaction between task and group ( $F(1,30) = 5.549$ ,  $p = .025$ ,  $\eta^2_p = 0.156$ ) and main effects of task ( $F(1,30) = 22.563$ ,  $p < .001$ ,  $\eta^2_p = 0.429$ ) and group ( $F(1,30) = 5.787$ ,  $p = .023$ ,  $\eta^2_p = 0.162$ ). Post-hoc independent  $t$ -tests revealed that there was a significant group difference in the voice identity task ( $t(30) = 3.228$ ,  $p = .003$ ,  $d = 1.141$ ) but not in the speech task ( $t(30) = 0.723$ ,  $p = .475$ ,  $d = 0.218$ ), indicating that the ASD group performed significantly worse than controls in voice identity but not in speech recognition.

##### *Vocal sound experiment*

Results from the vocal sound experiment are reported in the supplementary material (Supplementary Results; Supplementary Table 4; Supplementary Figure 5).

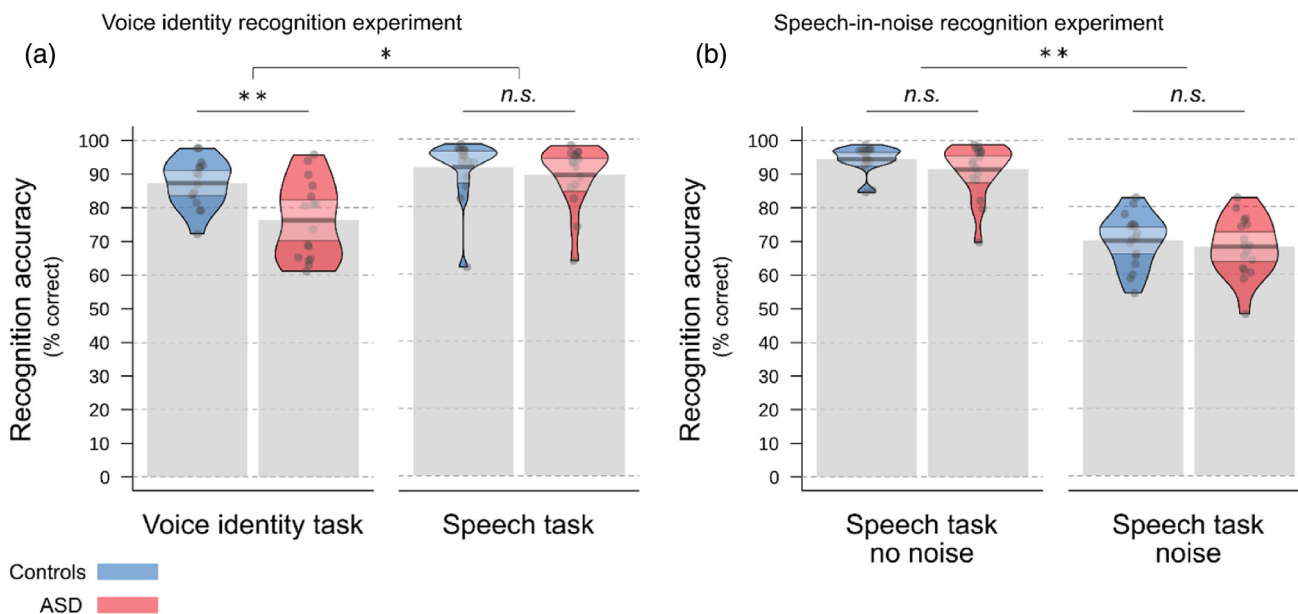
#### 4.1.2 | fMRI results

##### *Voice identity recognition experiment*

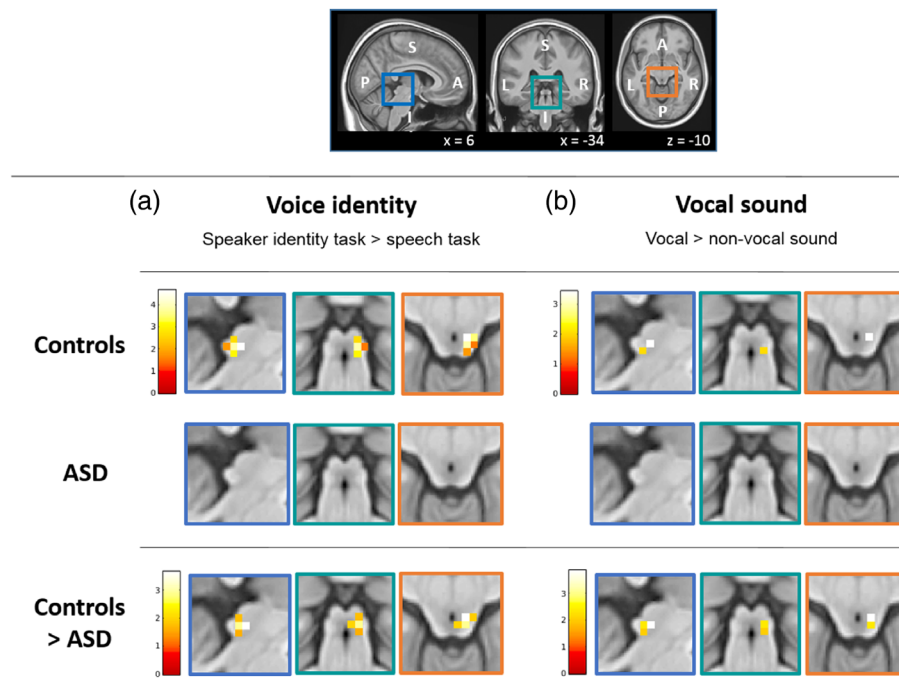
For the voice identity task as compared to the speech task, the control group had significantly higher BOLD response in the right IC ( $p = .002$  FWE corrected; Figure 4a). There was no such increased BOLD response for the voice identity task compared to the speech task in the IC within the ASD group (Table 2). The response difference between the voice identity and speech task was significantly higher for the control compared to the ASD group in the right IC ( $p = .006$  FWE corrected; Figure 4a; Supplementary Figure 6a). In the left IC and bilateral MGBs, we found no effects within or between the groups for the contrast “voice identity > speech” even at lenient statistical thresholds (all  $p$ s > .24 FWE corrected; Table 2).

##### *Vocal sound experiment*

When listening to vocal as compared to when listening to non-vocal sounds, the control group had higher BOLD responses in the right IC ( $p = .020$  FWE corrected; Figure 4b; Table 2); however, this difference did not survive the Bonferroni correction. There was no statistically significant difference between vocal and non-vocal sounds in the right IC responses within the ASD group. This was also not the case at a lenient statistical threshold ( $p > .65$  FWE corrected; Figure 4b; Table 2). There was a significant interaction between group (ASD/controls) and condition (vocal/non-vocal) in the right IC indicating that the difference between responses to vocal and non-vocal



**FIGURE 3** Performance accuracy in the voice identity recognition and the speech-in-noise recognition experiment. (a) In the voice identity recognition experiment, there was a significant interaction between group and task. The ASD group performed significantly worse than the control group in the voice identity task. There were no significant differences between the ASD and the control group in the speech task. (b) In the speech-in-noise recognition experiment, there were no significant group differences between the ASD and the control group. Both groups performed significantly worse in the noise condition as compared to the no noise speech task condition. Bars represent the mean average accuracy score for each group. Dots represent performances from each participant. Beans represent the smoothed density curve showing the full data distribution. Bands represent the confidence interval around the mean. \* $p < .05$ ; \*\* $p < .005$ ; n.s. not significant



**FIGURE 4** Functional magnetic resonance imaging (fMRI) results for the right inferior colliculus (IC) in the voice identity recognition experiment (a) and the vocal sound experiment (b). The control group showed enhanced blood oxygenation level-dependent (BOLD) responses in the right IC when performing the speaker identity task as compared to when performing the speech recognition task (a, *controls*) and when listening to vocal sounds as compared to when listening to non-vocal sounds (b, *controls*). There was no such enhanced response for the right IC within the ASD group (a and b, *ASD*) and this difference between the groups was significant (a and b, *controls* > *ASD*). Results are presented for the right IC and overlaid onto a group specific average image of normalised T1-weighted structural images. For ROI analyses in both experiments, we used independent masks provided in an atlas of the human subcortical auditory system (Sitek et al., 2019). The results are significant at  $p < .0125$  family wise error (FWE) corrected, and Bonferroni corrected for four ROIs (i.e., left and right IC and left and right medial geniculate body [MGB]). For display purposes only, the threshold of  $p = .05$  uncorrected was used. Colour bars represent  $t$ -values. P, posterior; S, superior; A, anterior; I, inferior; L, left; R, right; x, y, z, coordinates in Montreal Neurological Institute (MNI) space

sounds was higher in the control than in the ASD group ( $p = .003$  FWE corrected; Figure 4b; Table 2; Supplementary Figure 6b). In the left IC and bilateral MGBs, we found no statistically significant effects within or between the groups for the contrast “vocal > non-vocal sounds” (all  $p$ s > .056 FWE corrected; Table 2).

In both, the voice identity recognition and the vocal sound experiment, the results remained qualitatively the same when using ROIs based on previously reported coordinates (Supplementary Results; Supplementary Table 5).

## 4.2 | Speech-in-noise recognition experiment

### 4.2.1 | Behavioural results

Behavioural results were already reported and discussed previously (Schelinski & von Kriegstein, 2021). We here describe only the behavioural results that are relevant in the context of the present paper. Mean accuracy in the speech task was around 90% (no noise condition) and 70% (noise condition) (Supplementary Table 4). A repeated-measures ANOVA with the between-subject factor “group” (control,

ASD) and the within-subject factor “noise condition” (no noise, noise) revealed a significant main effect of noise condition ( $F(1,32) = 333.668, p < .001, \eta_p^2 = 0.912$ ) (Figure 3b). A post-hoc paired  $t$ -test indicated that over all participants, performance was lower in the noise as compared to the no noise condition ( $t(33) = 18.493, p < .001, d = 3.206$ ). There was no significant interaction between the factors task and group and no main effect of group ( $p$ s > .2).

### 4.2.2 | fMRI results

The control group showed higher BOLD responses in the left and right IC during the noise as compared to the no noise condition (left:  $p = .019$  FWE corrected, right:  $p = .025$  FWE corrected; Figure 5; Table 3). The results did, however, not survive Bonferroni correction for the four ROIs. In the ASD group, there were also higher responses elicited by the noise as compared to the no noise condition in the left IC ( $p = .048$  FWE corrected), but not for the right IC ( $p = .118$  FWE corrected). The group  $\times$  noise interaction was not significant (left:  $p = .129$  FWE corrected, right:  $p = .137$  FWE corrected; Table 3).

**TABLE 2** MNI-coordinates for significant BOLD-responses in the voice identity recognition experiment and the vocal sound experiment ( $p < .0125$  FWE corrected at peak level and Bonferroni corrected for four regions of interest). Italic coordinates indicate significant results at  $p < .05$  FWE corrected which did not survive Bonferroni correction. For information purposes only, grey font z-scores and  $p$ -values indicate results at FWE corrected thresholds ( $p > .05$  FWE corrected)

Voice identity task > speech task										
	Control group					ASD group				
	x	y	z	Z	p	x	y	z	Z	p
Right IC	6	-31	-10	3.61	.002	–			0.86	.537
Left IC				1.39	.321	–			0.19	.619
Right MGB				1.08	.439	–			0.15	.688
Left MGB				0.49	.690	–			-0.13	.745
Controls > ASD					ASD > controls					
Right IC	6	-31	-10	3.29	.006	–			0.10	.461
Left IC				1.33	.341	–			0.37	.663
Right MGB				1.52	.274	–			0.61	.596
Left MGB				1.71	.243	–			0.36	.719
Vocal sounds > non-vocal sounds										
	Control group					ASD group				
	x	y	z	Z	p	x	y	z	Z	p
Right IC	6	-31	-10	2.91	.020	–			0.45	.656
Left IC				1.98	.141	–			1.53	.274
Right MGB				2.25	.089	–			2.46	.056
Left MGB				1.89	.191	–			2.24	.100
Controls > ASD					ASD > controls					
Right IC	6	-31	-10	3.45	.003	–			0.66	.255
Left IC				2.10	.109	–			1.05	.459
Right MGB				0.64	.596	–			0.91	.510
Left MGB				0.47	.706	–			0.93	.567

Abbreviations: ASD, autism spectrum disorder; BOLD, blood oxygenation level-dependent; FWE, family wise error; IC, inferior colliculus; MGB, medial geniculate body; MNI, Montreal Neurological Institute; x, y, z, peak coordinates in MNI space (in mm).

In the bilateral MGBs we found no effects within or between the groups for the contrast “speech task noise > speech task no noise” even at lenient statistical thresholds (all  $ps > .136$  FWE corrected; Table 3).

The analyses with the ROIs provided in an independent atlas on the human subcortical auditory system (Sitek et al., 2019) confirmed the higher BOLD responses in the right IC in the control group ( $p = .011$  FWE corrected, Bonferroni corrected for the four ROIs), but not for the left IC ( $p = .177$  FWE corrected). The other results remained qualitatively the same (Supplementary Table 2).

### 4.3 | Control analyses

The control analyses showed that the results are unlikely due to unspecific group differences in head motion, task difficulty, or general BOLD response differences between groups (Supplementary Results; Supplementary Table 6; Supplementary Table 7). For the interested reader, whole-brain results from all three experiments can be found elsewhere (Schelinski et al., 2016; Schelinski & von Kriegstein, 2021).

## 5 | DISCUSSION

We showed that individuals with ASD had reduced subcortical sensory processing of voices within the IC – the central midbrain structure of the auditory pathway. There were three key findings:

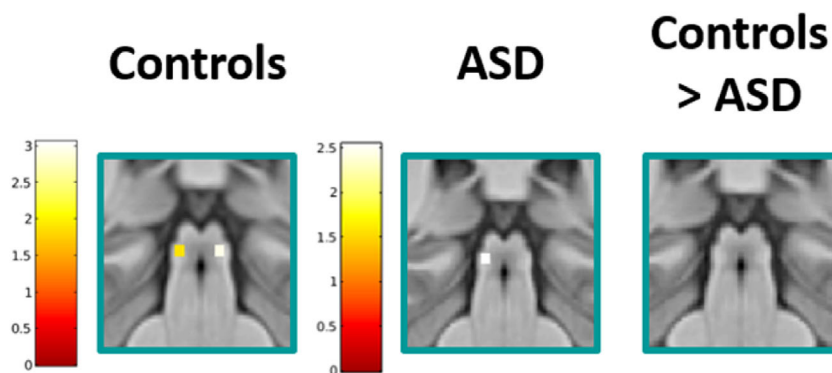
1. The right IC responded less in the ASD than in the control group for voice identity, in contrast to speech recognition.
2. The right IC also responded less in ASD than in controls when passively listening to vocal sounds in contrast to non-vocal sounds.
3. While in controls the left and right IC response was higher for recognising speech in background noise in contrast to clear speech, in ASD this was only the case for the left, but not the right IC.

Findings (1) and (2) survived extremely stringent statistical criteria and different ROI definitions. Together, the results reveal that impaired processing of communication signals in ASD is associated with altered responses in a specific subcortical auditory sensory pathway structure – the right IC.



## Speech-in-noise

Noise > no noise



**FIGURE 5** Functional magnetic resonance imaging (fMRI) results for the inferior colliculus (IC) in the speech-in-noise recognition experiment. For the control group, responses in the left and right IC were higher for the noise than for the no noise condition. For the same contrast, there were also higher responses in the left IC for the autism spectrum disorder (ASD) group (all  $p$ s < .05 family wise error [FWE] corrected). The results did, however, not survive Bonferroni correction for four ROIs (for  $p$  < .0125 FWE corrected). There was also no significant noise  $\times$  group interaction. Results are overlaid onto a group specific average image of normalised T1-weighted structural images. For ROI analyses, we used functionally defined independent masks created from the voice identity recognition experiment. For display purposes only, the threshold of  $p = .05$  uncorrected was used. Colour bars represent  $t$ -values. P, posterior; S, superior; A, anterior; I, inferior; L, left; R, right; x, y, z, coordinates in Montreal Neurological Institute (MNI) space

### Speech task noise > speech task no noise

	Control group					ASD group				
	x	y	z	Z	p	x	y	z	Z	p
Right IC	6	-34	-10	2.55	.025	—	—	—	1.19	.118
Left IC	-6	-31	-13	2.64	.019	-6	-34	-10	2.28	.048
Right MGB	—	—	—	0.15	.607	—	—	—	-0.15	.662
Left MGB	—	—	—	1.07	.377	—	—	—	0.16	.763
	Controls > ASD					ASD > controls				
Right IC	—	—	—	1.76	.137	—	—	—	0.80	.460
Left IC	—	—	—	1.75	.129	—	—	—	0.67	.469
Right MGB	—	—	—	1.68	.136	—	—	—	0.31	.577
Left MGB	—	—	—	1.47	.221	—	—	—	0.29	.716

Abbreviations: ASD, autism spectrum disorder; BOLD, blood oxygenation level-dependent; FWE, family wise error; IC, inferior colliculus; MGB, medial geniculate body; MNI, Montreal Neurological Institute; x, y, z, peak coordinates in MNI space (in mm).

**TABLE 3** MNI-coordinates for BOLD-responses in the speech-in-noise recognition experiment ( $p$  < .0125 FWE corrected at peak level and Bonferroni corrected for four regions of interest). *Italic coordinates indicate significant results at  $p$  < .05 FWE corrected which did not survive Bonferroni correction.* For information purposes only, grey font z-scores and  $p$ -values indicate results at FWE corrected thresholds ( $p$  > .05 FWE corrected).

The results are in agreement with an emerging field of research that suggests that altered sensory processing of communication signals is an integral part in explaining difficulties in social cognition in ASD (reviewed in Baum et al., 2015; Robertson & Baron-Cohen, 2017; Thye et al., 2018). The notion that ASD is associated with sensory dysfunction is a long-standing topic. Classic theories on ASD, such as

the weak central coherence theory, describe a tendency to focus on the detail, while the ability to integrate elements into a coherent percept is weak (reviewed in Haesen, Boets, & Wagemans, 2011; Happe & Frith, 2006). There are also long-standing suggestions that the motion perception deficits in ASD could be linked to the social difficulties observed in ASD (Dakin & Frith, 2005). However, to-date, this

topic is mostly covered as sensory difficulties accompanying ASD. Sensory contributions to ASD symptomatology and impairments are often focused on hyper- and hypo-sensory processing which usually refers to an enhanced ability to perceive sensory stimuli or absent or less response to sensory input (DSM-5, APA, 2013; reviewed in Pellicano, 2013; Robertson & Baron-Cohen, 2017). The relation of these sensory difficulties to social cognition is largely unknown. We suggest that sensory processing difficulties are important in explaining difficulties in social cognition in ASD and that these sensory difficulties are associated with subcortical sensory dysfunction. Whether sensory processing difficulties cause or are a consequence of social cognition difficulties and its relation to atypical sensory responses remains an open question. However, studies on animal models of ASD suggest that structural alteration of the IC might already occur in utero (Zimmerman, Smith, Fech, Mansour, & Kulesza Jr., 2020) and that brainstem alterations might influence those brain structures, such as the cerebral cortex, that develop later (reviewed in Dadalco & Travers, 2018).

One very relevant open question is whether the subcortical sensory processing dysfunction is a downstream effect of dysfunctional cerebral cortex or whether it is the other way around (i.e., reflect a bottom-up sensory processing deficit). Here, the same ASD sample which showed – in comparison to controls – reduced right STS/G responses when performing a voice identity recognition task as compared to when performing a speech recognition task (Schelinski et al., 2016) showed reduced right IC responses for the same contrast. Additionally, the same ASD group who showed a lack in enhanced right IC responses when performing the speech-in-noise task as compared to the speech task without noise in the current study also showed reduced cerebral cortex responses in the left inferior frontal gyrus for the same contrast (Schelinski & von Kriegstein, 2021). These results might indicate that altered processing of voices in ASD that can be observed at the cerebral cortex level is already present in the IC. Whether the reduced cerebral cortex responses are the result of reduced subcortical sensory pathway responses or vice versa is difficult to assess with the low temporal resolution of fMRI and given the vast amount of feedback connections from cerebral cortex to the IC (Coomes Peterson & Schofield, 2007; Javad et al., 2014; Suga, Gao, Zhang, Ma, & Olsen, 2000; Winer, Miller, Lee, & Schreiner, 2005). For the control group, the enhanced IC-responses when performing a task might indicate a top-down modulation, which is less present or altered in the ASD group. In this context, it is interesting that there was reduced right IC response in the ASD in contrast to the control group for vocal sounds (in contrast to non-vocal sounds) even in a passive listening design. Furthermore, in this design there were no differences between groups on the level of the voice-sensitive cerebral cortex areas (i.e., right STS/G) when listening to vocal as compared to non-vocal sounds (Schelinski et al., 2016). This could indicate that it is the right IC dysfunction that is critical for the behavioural difficulties in voice perception, while the reduced cerebral cortex responses when performing tasks might be a downstream effect. The IC shares a rich connectivity with other subcortical and cortical brain structures (e.g., Huffman & Henson, 1990; Sitek et al., 2019; Stebbings,

Lesicko, & Llano, 2014; Winer, 2006). Potentially, dysfunctional processing at this stage might be compensated by interacting with other brain regions (compare Glick & Sharma, 2017; Lopez-Barroso & de Diego-Balaguer, 2017; Occelli, Spence, & Zampini, 2013; Roswadowitz et al., 2017) thereby explaining why cerebral cortex responses to passive listening of voices were on a typical level while the right IC was atypical.

It was fascinating that it was the IC that showed the reduced responses to voices in ASD. Brainstem responses have been found to be altered in children with ASD as compared to typically developing children when passively listening to syllables when the speech signal showed differences in pitch contour (Russo et al., 2008) and when the speech signal was presented with and without additional white noise (Russo et al., 2009). Conversely, the ASD groups had typical brainstem responses when listening to non-vocal sounds; that is, click sounds (Russo et al., 2008; Russo et al., 2009). The critical role of the IC in decoding and integrating spectro-temporal sound information including frequency processing has been shown in a large body of animal and human studies (e.g., Baumann et al., 2011; Chandrasekaran, Kraus, & Wong, 2012; Griffiths et al., 2001; Moerel et al., 2015; reviewed in Krishnan & Gandour, 2009; Pannese, Grandjean, & Fruhholz, 2015). Potentially, altered processing of voice acoustic features in the IC might be associated with impaired processing of auditory communication signals in ASD. The perception and integration of a variety of acoustic voice features, such as the fundamental frequency (F0), is crucial for recognising voice identity (Baumann & Belin, 2010; Gaudrain, Li, Ban, & Patterson, 2009; Kreitewolf et al., 2014) and speech-in-noise (reviewed in Anderson & Kraus, 2010; Brown & Bacon, 2010). F0 is the lowest carrier frequency of the voice signal and is perceived as vocal pitch. In ASD, differences in processing voice acoustic features, such as vocal pitch, might be crucial in explaining difficulties in voice recognition including difficulties in voice identity (Schelinski et al., 2016; Schelinski et al., 2017) and speech-in-noise recognition (Schelinski & von Kriegstein, 2020). The same ASD group who participated in the voice identity and the vocal sound experiment showed impaired vocal, but intact non-vocal pitch perception in a previous study (Schelinski et al., 2017; also see Jiang, Liu, Wan, & Jiang, 2015). Although there are dedicated pitch perception areas in the cerebral cortex (De Angelis et al., 2018; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Puschmann, Uppenkamp, Kollmeier, & Thiel, 2010), the IC has also been associated with the representation of pitch (Bianchi et al., 2017; Chandrasekaran et al., 2012; Griffiths et al., 2001; reviewed in Gruters & Groh, 2012; Pannese et al., 2015). Thus, the IC might be a candidate structure for explaining altered vocal pitch perception in ASD. Critically, for speech-in-noise recognition our results only provide first evidence for IC dysfunction in ASD, since there was no significant interaction between group (ASD, controls) and the noise condition (noise, no noise). Our research opens up the research questions in how far impaired vocal pitch processing in ASD can be attributed to altered IC functioning.

The finding that it is the IC and not the MGB that shows significant differences between groups is fascinating as it may reveal a

specific deficit associated with the function of the IC. We speculate that the IC rather than the MGB is associated with the processing of those voice features, which are important for vocal pitch processing (for reviews, see Gruters & Groh, 2012; Pannese et al., 2015). Our assumption is in line with previous results in which we found worse performance in ASD in a task on vocal pitch discrimination as compared to typically developed controls (Schelinski et al., 2017, also see Jiang et al., 2015). This deficit in vocal pitch perception was relatively selective; that is, the same ASD group participants did perform equally well as the control group in tasks on vocal timbre and non-vocal pitch perception (Schelinski et al., 2017). However, our assumption of a specific relation between altered vocal pitch processing in ASD and altered IC processing remains speculative.

Besides voice identity and speech-in-noise recognition difficulties, ASD has been associated with impaired vocal emotion recognition (Globerson, Amir, Kishon-Rabin, & Golan, 2015; Golan, Baron-Cohen, Hill, & Rutherford, 2007; Rutherford, Baron-Cohen, & Wheelwright, 2002; reviewed in Lartseva, Dijkstra, & Buitelaar, 2014). As for voice identity and speech-in-noise recognition (e.g., Anderson & Kraus, 2010; Gaudrain et al., 2009), F0 is important for processing vocal emotion (Fairbanks & Pronovost, 1938; Gold et al., 2012; Quam & Swingle, 2012). Difficulties in processing vocal pitch (i.e., discrimination of differences in F0) in ASD might also be related to impaired vocal emotion recognition (Schelinski & von Kriegstein, 2019). The IC might be one of the key structures that provide a first acoustic profile of vocal emotions based on spectral-temporal features of the voice signal which is essential for further differentiation of vocal expressions (Pannese et al., 2015). In line with this view and previous results in ASD, we speculate that differences in IC responses might be critical for explaining vocal emotion recognition difficulties in ASD.

The role of sensory processing impairments as contributors to difficulties in social cognition is also discussed in schizophrenia (Javitt & Freedman, 2015) – a disorder that shares genetic and behavioural characteristics with ASD (Chisholm, Lin, Abu-Akel, & Wood, 2015; Owen & O'Donovan, 2017). In ASD and in schizophrenia, voice perception difficulties are related to altered pitch perception (Globerson et al., 2015; Gold et al., 2012; Jahshan, Wynn, & Green, 2013; Kantrowitz et al., 2013; Leitman et al., 2011; Leitman et al., 2010; Schelinski & von Kriegstein, 2019; but also see Chhabra, Badcock, Maybery, & Leung, 2012). In this context, it is interesting that in schizophrenia there are also first indications of reduced functioning of the IC (Gaebler et al., 2020). A transdiagnostic approach is important, since it holds the potential to further improve the characterisation and treatment of different forms of psychopathology (Kapur, Phillips, & Insel, 2012), an often challenging task in clinical practice.

## 6 | LIMITATIONS

Our study rests on three different experiments and different participant samples ( $n = 32$  in the voice identity recognition experiment,  $n = 32$  in the vocal sound experiment and  $n = 34$  in the speech-in-

noise experiment). The samples were, however, relatively homogeneous (i.e., adults with an IQ and verbal abilities at least within the normal range). This means that we do not know whether similar alterations as found in the present study can also be found in a more heterogeneous sample representing the whole autism spectrum. The present study pioneers investigation of subcortical sensory structures with relatively high spatial resolution in ASD. The spatial resolution could be increased even more by not using whole brain data, as in the present study, but by focusing only on the subcortical sensory pathway structures. We did not do this in the present study as data were available that we acquired previously to investigate voice processing at the cerebral cortex level (Schelinski et al., 2016; Schelinski & von Kriegstein, 2021). Acquiring such neuroimaging data in special populations is extremely time consuming and cost-intensive.

## 7 | IMPLICATIONS AND OUTLOOK

Voice processing is an evolutionary preserved process (Petkov, Logothetis, & Obleser, 2009). Voice-specific responses are already present in utero (Kisilevsky et al., 2003) and voice-specific brain responses develop early in human life (Grossmann, Oberecker, Koch, & Friederici, 2010). In ASD, altered voice perception such as a lack of preference for the mother's voice can already be observed in early infancy (Klin, 1991). Alterations in auditory subcortical sensory processing might at least partly contribute to the development of difficulties in auditory communication. Although the investigation of developmental changes in subcortical sensory pathway structures in the typically and atypically developing brain remains a subject to study, a more mechanistic understanding of these processes is likely a good basis for diagnostic markers. For example, testing impairments in voice identity, vocal emotion recognition or perception of acoustic voice features such as vocal pitch, might be a straightforward additional tool in the diagnostic procedure of ASD. It might also be a good basis for evaluating therapeutic options since there is evidence for plasticity in subcortical auditory processing (reviewed in Chandrasekaran, Skoe, & Kraus, 2014). Alterations in the sensory processing of communication signals in ASD might have the potential to facilitate describing and treating ASD social communication symptoms in the future.

### ACKNOWLEDGEMENTS

The authors are grateful to our participants for taking part in the study. The authors thank the teams of the IT and the graphics department of the Max Planck Institute for Human Cognitive and Brain Sciences for technical support in creating and processing stimulus material. Open access funding enabled and organized by Projekt DEAL.

### AUTHOR CONTRIBUTIONS

**Stefanie Schelinski** and **Katharina von Kriegstein**: Conceptualised the study and designed the methodology. **Stefanie Schelinski**: Managed and coordinated the study and performed the experiments. **Stefanie Schelinski** and **Alejandro Tabas**: Analysed and visualised the data.

**Stefanie Schelinski and Katharina von Kriegstein:** Wrote the original manuscript. **Alejandro Tabas:** Critically reviewed the initial draft. **Katharina von Kriegstein:** Supervised and acquired funding for the study. All authors read and approved the final manuscript.

## CONFLICT OF INTERESTS

The authors declare no potential conflict of interests.

## ETHICS STATEMENT AND CONSENT TO PARTICIPATE

The studies were approved by the Ethics Committee of the Medical Faculty at the University Leipzig, Germany (299-12-14092012). All participants gave written informed consent in accordance with the Declaration of Helsinki and procedures approved by the Research Ethics Committee of the University of Leipzig.

## DATA AVAILABILITY STATEMENT

The datasets generated and/or analysed during the current study are not publicly available as we do not have consent from all participants to do this. SPM files are available from the corresponding author on request.

## ORCID

Stefanie Schelinski  <https://orcid.org/0000-0001-5398-0020>

Alejandro Tabas  <https://orcid.org/0000-0002-8643-1543>

Katharina von Kriegstein  <https://orcid.org/0000-0001-7989-5860>

## REFERENCES

- Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., ... Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29), 12060–12065. <https://doi.org/10.1073/pnas.1302982110>
- Abrams, D. A., Padmanabhan, A., Chen, T., Odriozola, P., Baker, A. E., Kochalka, J., ... Menon, V. (2019). Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *eLife*, 8, e39906. <https://doi.org/10.7554/eLife.39906>
- Alcantara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45(6), 1107–1114. <https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x>
- Amaral, L., Ganho-Avila, A., Osorio, A., Soares, M. J., He, D., Chen, Q., ... Almeida, J. (2016). Hemispheric asymmetries in subcortical visual and auditory relay structures in congenital deafness. *European Journal of Neuroscience*, 44(6), 2334–2339. <https://doi.org/10.1111/ejn.13340>
- American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th ed.). Washington, DC: American Psychiatric Association.
- 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. <https://doi.org/10.3766/jaaa.21.9.3>
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17. <https://doi.org/10.1023/a:1005653411471>
- 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. <https://doi.org/10.1016/j.pneurobio.2015.09.007>
- Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research*, 74(1), 110–120. <https://doi.org/10.1007/s00426-008-0185-z>
- Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., & Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nature Neuroscience*, 14(4), 423–425. <https://doi.org/10.1038/nn.2771>
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135. <https://doi.org/10.1016/j.tics.2004.01.008>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312. <https://doi.org/10.1038/35002078>
- Bianchi, F., Hjortkjaer, J., Santurette, S., Zatorre, R. J., Siebner, H. R., & Dau, T. (2017). Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. *NeuroImage*, 163, 398–412. <https://doi.org/10.1016/j.neuroimage.2017.07.057>
- Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage*, 175, 56–69. <https://doi.org/10.1016/j.neuroimage.2018.03.060>
- Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience and Biobehavioral Reviews*, 47, 717–734. <https://doi.org/10.1016/j.neubiorev.2014.10.022>
- Bölte, S., & Poustka, F. (2006). *Fragebogen zur Sozialen Kommunikation (FSK)*. Bern: Verlag Hans Huber.
- Bölte, S., Rühl, D., Schmötzer, G., & Poustka, F. (2003). *Diagnostisches Interview für Autismus - Revidiert (ADI-R)*. Bern: Verlag Hans Huber.
- Bonilha, L., Hillis, A. E., Hickok, G., den Ouden, D. B., Rorden, C., & Fridriksson, J. (2017). Temporal lobe networks supporting the comprehension of spoken words. *Brain*, 140(9), 2370–2380. <https://doi.org/10.1093/brain/awx169>
- Borowiak, K., Schelinski, S., & von Kriegstein, K. (2018). Recognizing visual speech: Reduced responses in visual-movement regions, but not other speech regions in autism. *NeuroImage: Clinical*, 20, 1078–1091. <https://doi.org/10.1016/j.nicl.2018.09.019>
- Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 39(2), 171–181.
- Brickenkamp, R. (2002). *Test d2 - Aufmerksamkeits-Belastung-Test (d2)*. Göttingen: Hogrefe.
- Brown, C. A., & Bacon, S. P. (2010). Fundamental frequency and speech intelligibility in background noise. *Hearing Research*, 266(1–2), 52–59. <https://doi.org/10.1016/j.heares.2009.08.011>
- Chandrasekaran, B., Kraus, N., & Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, 107(5), 1325–1336. <https://doi.org/10.1152/jn.00923.2011>
- Chandrasekaran, B., Skoe, E., & Kraus, N. (2014). An integrative model of subcortical auditory plasticity. *Brain Topography*, 27(4), 539–552. <https://doi.org/10.1007/s10548-013-0323-9>
- Chhabra, S., Badcock, J. C., Maybery, M. T., & Leung, D. (2012). Voice identity discrimination in schizophrenia. *Neuropsychologia*, 50(12), 2730–2735. <https://doi.org/10.1016/j.neuropsychologia.2012.08.006>
- Chisholm, K., Lin, A., Abu-Akel, A., & Wood, S. J. (2015). The association between autism and schizophrenia spectrum disorders: A review of eight alternate models of co-occurrence. *Neuroscience and Biobehavioral Reviews*, 55, 173–183. <https://doi.org/10.1016/j.neubiorev.2015.04.012>
- Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., ... Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, 10(1), 5036. <https://doi.org/10.1038/s41467-019-13003-w>
- Coomes Peterson, D., & Schofield, B. R. (2007). Projections from auditory cortex contact ascending pathways that originate in the superior olive and inferior colliculus. *Hearing Research*, 232(1–2), 67–77. <https://doi.org/10.1016/j.heares.2007.06.009>

- Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R., 3rd. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences*, 23(2), 114–127. <https://doi.org/10.1016/j.tics.2018.12.001>
- Dadalko, O. I., & Travers, B. G. (2018). Evidence for brainstem contributions to autism spectrum disorders. *Frontiers in Integrative Neuroscience*, 12, 47. <https://doi.org/10.3389/fnint.2018.00047>
- Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, 48(3), 497–507. <https://doi.org/10.1016/j.neuron.2005.10.018>
- De Angelis, V., De Martino, F., Moerel, M., Santoro, R., Hausfeld, L., & Formisano, E. (2018). Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds. *NeuroImage*, 180(Pt A), 291–300. <https://doi.org/10.1016/j.neuroimage.2017.11.020>
- Diaz, B., Hintz, F., Kiebel, S. J., & von Kriegstein, K. (2012). Dysfunction of the auditory thalamus in developmental dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34), 13841–13846. <https://doi.org/10.1073/pnas.1119828109>
- Fairbanks, G., & Pronovost, W. (1938). Vocal pitch during simulated emotion. *Science*, 88(2286), 382–383. <https://doi.org/10.1126/science.88.2286.382>
- Forstmann, B. U., de Hollander, G., van Maanen, L., Alkemade, A., & Keuken, M. C. (2016). Towards a mechanistic understanding of the human subcortex. *Nature Reviews Neuroscience*, 18(1), 57–65. <https://doi.org/10.1038/nrn.2016.163>
- Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., ... von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(4), 280–289.
- Friston, K., Ashburner, A., Kiebel, S., Nichols, T., & Penny, W. (2007). *Statistical parametric mapping: The analysis of functional brain images*. London: Academic Press.
- Gaebler, A. J., Zweerings, J., Koten, J. W., Konig, A. A., Turetsky, B. I., Zvyagintsev, M., & Mathiak, K. (2020). Impaired subcortical detection of auditory changes in schizophrenia but not in major depression. *Schizophrenia Bulletin*, 46(1), 193–201. <https://doi.org/10.1093/schbul/sbz027>
- Gaudrain, E., Li, S., Ban, V. S., & Patterson, R. D. (2009). *The role of glottal pulse rate and vocal tract length in the perception of speaker identity*. Interspeech 2009: 10th Annual Conference of the International Speech Communication Association 2009.
- Gervais, H., Belin, P., Boddart, N., Leboyer, M., Coez, A., Sfaello, I., ... Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, 7(8), 801–802. <https://doi.org/10.1038/nn1291>
- Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing Research*, 343, 191–201. <https://doi.org/10.1016/j.heares.2016.08.012>
- Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, 8(2), 153–163. <https://doi.org/10.1002/aur.1432>
- Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'reading the mind in the voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, 37(6), 1096–1106. <https://doi.org/10.1007/s10803-006-0252-5>
- Gold, R., Butler, P., Revheim, N., Leitman, D. I., Hansen, J. A., Gur, R. C., ... Javitt, D. C. (2012). Auditory emotion recognition impairments in schizophrenia: Relationship to acoustic features and cognition. *The American Journal of Psychiatry*, 169(4), 424–432. <https://doi.org/10.1176/appi.ajp.2011.11081230>
- Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience*, 4(6), 633–637. <https://doi.org/10.1038/88459>
- Groen, W. B., van Orsouw, L., 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. <https://doi.org/10.1007/s10803-008-0682-3>
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–858. <https://doi.org/10.1016/j.neuron.2010.03.001>
- Gruters, K. G., & Groh, J. M. (2012). Sounds and beyond: Multisensory and other non-auditory signals in the inferior colliculus. *Frontiers in Neural Circuits*, 6, 96. <https://doi.org/10.3389/fncir.2012.00096>
- Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5(2), 701–714. <https://doi.org/10.1016/j.rasd.2010.11.006>
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., ... Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3), 213–223. [https://doi.org/10.1002/\(sici\)1097-0193\(1999\)7:3<213::aid-hbm5>3.0.co;2-n](https://doi.org/10.1002/(sici)1097-0193(1999)7:3<213::aid-hbm5>3.0.co;2-n)
- Happe, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5–25. <https://doi.org/10.1007/s10803-005-0039-0>
- Herrington, J. D., Baron-Cohen, S., Wheelwright, S. J., Singh, K. D., Bullmore, E. T., Brammer, M., & Williams, S. C. R. (2007). The role of MT+/V5 during biological motion perception in Asperger syndrome: An fMRI study. *Research in Autism Spectrum Disorders*, 1(1), 14–27. <https://doi.org/10.1016/j.rasd.2006.07.002>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Huffman, R. F., & Henson, O. W., Jr. (1990). The descending auditory pathway and acousticomotor systems: Connections with the inferior colliculus. *Brain Research Reviews*, 15(3), 295–323. [https://doi.org/10.1016/0165-0173\(90\)90005-9](https://doi.org/10.1016/0165-0173(90)90005-9)
- Jahshan, C., Wynn, J. K., & Green, M. F. (2013). Relationship between auditory processing and affective prosody in schizophrenia. *Schizophrenia Research*, 143(2–3), 348–353. <https://doi.org/10.1016/j.schres.2012.11.025>
- Javad, F., Warren, J. D., Micallef, C., Thornton, J. S., Golay, X., Yousry, T., & Mancini, L. (2014). Auditory tracts identified with combined fMRI and diffusion tractography. *NeuroImage*, 84, 562–574. <https://doi.org/10.1016/j.neuroimage.2013.09.007>
- Javitt, D. C., & Freedman, R. (2015). Sensory processing dysfunction in the personal experience and neuronal machinery of schizophrenia. *The American Journal of Psychiatry*, 172(1), 17–31. <https://doi.org/10.1176/appi.ajp.2014.13121691>
- Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, 34(1), 65–73. <https://doi.org/10.1002/mrm.1910340111>
- Jiang, J., Liu, F., Wan, X., & Jiang, C. (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. <https://doi.org/10.1007/s10803-015-2370-4>
- Kang, D. H., Kwon, K. W., Gu, B. M., Choi, J. S., Jang, J. H., & Kwon, J. S. (2008). Structural abnormalities of the right inferior colliculus in schizophrenia. *Psychiatry Research*, 164(2), 160–165. <https://doi.org/10.1016/j.psychres.2007.12.023>
- Kantrowitz, J. T., Leitman, D. I., Lehrfeld, J. M., Laukka, P., Juslin, P. N., Butler, P. D., ... Javitt, D. C. (2013). Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder. *Schizophrenia Bulletin*, 39(1), 86–93. <https://doi.org/10.1093/schbul/sbr060>
- Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17(12), 1174–1179. <https://doi.org/10.1038/mp.2012.105>



- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., ... Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, 14(3), 220–224. <https://doi.org/10.1111/1467-9280.02435>
- 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. <https://doi.org/10.3389/fpsyg.2013.00578>
- Klin, A. (1991). Young autistic childrens listening preferences in regard to speech—A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, 21(1), 29–42. <https://doi.org/10.1007/Bf02206995>
- Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage*, 91, 375–385. <https://doi.org/10.1016/j.neuroimage.2014.01.005>
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, 110(3), 135–148. <https://doi.org/10.1016/j.bandl.2009.03.005>
- Kummerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Kloppel, S., ... Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain*, 136(Pt 2), 619–629. <https://doi.org/10.1093/brain/aws354>
- Larteva, A., Dijkstra, T., & Buitelaar, J. K. (2014). Emotional language processing in autism spectrum disorders: A systematic review. *Frontiers in Human Neuroscience*, 8, 991. <https://doi.org/10.3389/fnhum.2014.00991>
- Leff, A. P., Schofield, T. M., Crinion, J. T., Seghier, M. L., Grogan, A., Green, D. W., & Price, C. J. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain*, 132, 3401–3410. <https://doi.org/10.1093/brain/awp273>
- Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P., & Javitt, D. C. (2010). Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin*, 36(3), 545–556. <https://doi.org/10.1093/schbul/sbn115>
- Leitman, D. I., Wolf, D. H., Laukka, P., Ragland, J. D., Valdez, J. N., Turetsky, B. I., ... Gur, R. C. (2011). Not pitch perfect: Sensory contributions to affective communication impairment in schizophrenia. *Biological Psychiatry*, 70(7), 611–618.
- Lopez-Barroso, D., & de Diego-Balaguer, R. (2017). Language learning variability within the dorsal and ventral streams as a cue for compensatory mechanisms in aphasia recovery. *Frontiers in Human Neuroscience*, 11, 476. <https://doi.org/10.3389/fnhum.2017.00476>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., 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. <https://doi.org/10.1023/A:1005592401947>
- 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. <https://doi.org/10.1007/BF02172145>
- Maguinness, C., Roswadowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*, 116(Pt B), 179–193. <https://doi.org/10.1016/j.neuropsychologia.2018.03.039>
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- Moerel, M., De Martino, F., Ugurbil, K., Yacoub, E., & Formisano, E. (2015). Processing of frequency and location in human subcortical auditory structures. *Scientific Reports*, 5, 17048. <https://doi.org/10.1038/srep17048>
- Mugler, J. P., 3rd, & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15(1), 152–157. <https://doi.org/10.1002/mrm.1910150117>
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376. <https://doi.org/10.3758/bf03206860>
- Occelli, V., Spence, C., & Zampini, M. (2013). Auditory, tactile, and audiotactile information processing following visual deprivation. *Psychological Bulletin*, 139(1), 189–212. <https://doi.org/10.1037/a0028416>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Owen, M. J., & O'Donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum: Evidence from genomics. *World Psychiatry*, 16(3), 227–235. <https://doi.org/10.1002/wps.20440>
- Pannese, A., Grandjean, D., & Fruhholz, S. (2015). Subcortical processing in auditory communication. *Hearing Research*, 328, 67–77. <https://doi.org/10.1016/j.heares.2015.07.003>
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776. [https://doi.org/10.1016/s0896-6273\(02\)01060-7](https://doi.org/10.1016/s0896-6273(02)01060-7)
- Pellicano, E. (2013). Sensory symptoms in autism: A blooming, buzzing confusion? *Child Development Perspectives*, 7(3), 143–148. <https://doi.org/10.1111/cdep.12031>
- Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *The Neuroscientist*, 15(5), 419–429. <https://doi.org/10.1177/1073858408326430>
- Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology*, 40(5), 221–244.
- Poliva, O., Bestelmeyer, P. E., Hall, M., Bultitude, J. H., Koller, K., & Rafal, R. D. (2015). Functional mapping of the human auditory cortex: fMRI investigation of a patient with auditory agnosia from trauma to the inferior colliculus. *Cognitive and Behavioral Neurology*, 28(3), 160–180. <https://doi.org/10.1097/WNN.0000000000000072>
- Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, 49(2), 1641–1649. <https://doi.org/10.1016/j.neuroimage.2009.09.045>
- Quam, C., & Swingle, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, 83(1), 236–250. <https://doi.org/10.1111/j.1467-8624.2011.01700.x>
- R Core Team. (2021). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Retrieved from <https://www.R-project.org/>
- Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18(11), 671–684. <https://doi.org/10.1038/nrn.2017.112>
- Roswadowitz, C., Kappes, C., Obrig, H., & von Kriegstein, K. (2018). Obligatory and facultative brain regions for voice-identity recognition. *Brain*, 141(1), 234–247. <https://doi.org/10.1093/brain/awx313>
- Roswadowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology*, 24(19), 2348–2353. <https://doi.org/10.1016/j.cub.2014.08.048>
- Roswadowitz, C., Schelinski, S., & von Kriegstein, K. (2017). Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. *NeuroImage*, 155, 97–112. <https://doi.org/10.1016/j.neuroimage.2017.02.064>

- Rühl, D., Bölte, S., Feineis-Matthews, S., & Poustka, F. (2004). *Diagnostische Beobachtungsskala für Autistische Störungen (ADOS)*. Bern: Verlag Hans Huber.
- Russo, N., Nicol, T., Trommer, B., Zecker, S., & Kraus, N. (2009). Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Developmental Science*, 12(4), 557–567. <https://doi.org/10.1111/j.1467-7687.2008.00790.x>
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119(8), 1720–1731. <https://doi.org/10.1016/j.clinph.2008.01.108>
- Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, 32(3), 189–194. <https://doi.org/10.1023/a:1015497629971>
- Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire (SCQ)*. Los Angeles, CA: Western Psychological Services.
- 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. <https://doi.org/10.1093/scan/nsw089>
- 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. <https://doi.org/10.1016/j.neuropsychologia.2014.09.031>
- Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research*, 10(1), 155–168. <https://doi.org/10.1002/aur.1639>
- 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. <https://doi.org/10.1007/s10803-018-3681-z>
- Schelinski, S., & von Kriegstein, K. (2020). Brief report: Speech-in-noise recognition and the relation to vocal pitch perception in adults with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*, 50(1), 356–363. <https://doi.org/10.1007/s10803-019-04244-1>
- Schelinski, S. & von Kriegstein, K. (2021). Reduced neural processing of speech-in-noise in the left inferior frontal gyrus in autism spectrum disorder. *PsyArXiv*. <https://doi.org/10.31234/osf.io/rf8tp>
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24–30. <https://doi.org/10.1016/j.tics.2005.11.009>
- Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, 366(6461), 58–62. <https://doi.org/10.1126/science.aax0288>
- Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, 385(6613), 254–257. <https://doi.org/10.1038/385254a0>
- Sheppard, S. M., Keator, L. M., Breining, B. L., Wright, A. E., Saxena, S., Tippett, D. C., & Hillis, A. E. (2020). Right hemisphere ventral stream for emotional prosody identification: Evidence from acute stroke. *Neurology*, 94(10), E1013–E1020. <https://doi.org/10.1212/Wnl.0000000000009052>
- Sitek, K. R., Gulban, O. F., Calabrese, E., Johnson, G. A., Lage-Castellanos, A., Moerel, M., ... De Martino, F. (2019). Mapping the human subcortical auditory system using histology, postmortem MRI and in vivo MRI at 7T. *eLife*, 8, e48932. <https://doi.org/10.7554/eLife.48932>
- Stebbing, K. A., Lesicko, A. M., & Llano, D. A. (2014). The auditory corticocollicular system: Molecular and circuit-level considerations. *Hearing Research*, 314, 51–59. <https://doi.org/10.1016/j.heares.2014.05.004>
- Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, J. F. (2000). The corticofugal system for hearing: Recent progress. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11807–11814. <https://doi.org/10.1073/pnas.97.22.11807>
- Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, 137(4), 682–707. <https://doi.org/10.1037/a0023987>
- Tabas, A., Mihai, G., Kiebel, S., Trampel, R., & von Kriegstein, K. (2020). Abstract rules drive adaptation in the subcortical sensory pathway. *eLife*, 9, e64501. <https://doi.org/10.7554/eLife.64501>
- Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., & Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Developmental Cognitive Neuroscience*, 29, 151–167. <https://doi.org/10.1016/j.dcn.2017.04.010>
- 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. <https://doi.org/10.1007/s40489-017-0111-7>
- Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, 1(2), 185–195. [https://doi.org/10.1016/0278-2626\(82\)90016-1](https://doi.org/10.1016/0278-2626(82)90016-1)
- von Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene (WIE)*. Frankfurt: Harcourt Test Services.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, 17(1), 48–55. [https://doi.org/10.1016/S0926-6410\(03\)00079-X](https://doi.org/10.1016/S0926-6410(03)00079-X)
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS-III)*. San Antonio, TX: The Psychological Corporation.
- White, T., van der Ende, J., & Nichols, T. E. (2019). Beyond Bonferroni revisited: Concerns over inflated false positive research findings in the fields of conservation genetics, biology, and medicine. *Conservation Genetics*, 20(4), 927–937. <https://doi.org/10.1007/s10592-019-01178-0>
- Whitehead, J. C., & Armony, J. L. (2018). Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Human Brain Mapping*, 39(12), 4913–4924. <https://doi.org/10.1002/hbm.24333>
- Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hearing Research*, 212(1–2), 1–8. <https://doi.org/10.1016/j.heares.2005.06.014>
- Winer, J. A., Miller, L. M., Lee, C. C., & Schreiner, C. E. (2005). Auditory thalamocortical transformation: Structure and function. *Trends in Neurosciences*, 28(5), 255–263. <https://doi.org/10.1016/j.tins.2005.03.009>
- World Health Organization. (2004). *International statistical classification of diseases and related health problems-10th revision* (2nd ed.). Geneva.
- Zimmerman, R., Smith, A., Fech, T., Mansour, Y., & Kulesza, R. J., Jr. (2020). In utero exposure to valproic acid disrupts ascending projections to the central nucleus of the inferior colliculus from the auditory brainstem. *Experimental Brain Research*, 238(3), 551–563. <https://doi.org/10.1007/s00221-020-05729-7>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Schelinski, S., Tabas, A., & von Kriegstein, K. (2022). Altered processing of communication signals in the subcortical auditory sensory pathway in autism. *Human Brain Mapping*, 43(6), 1955–1972. <https://doi.org/10.1002/hbm.25766>