Mechanisms of Voice Processing:

Evidence from Autism Spectrum Disorder

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

eingereicht an der

Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Dipl.-Psych. Stefanie Schelinski

Tag der Verteidigung: 18.12.2017

Abstract

The correct perception of information carried by the voice is a key requirement for successful human communication. Hearing another person's voice provides information about who is speaking (voice identity), what is said (vocal speech) and the emotional state of a person (vocal emotion). Autism spectrum disorder (ASD) is associated with impaired voice identity and vocal emotion perception while the perception of vocal speech is relatively intact. However, the underlying mechanisms of these voice perception impairments are unknown. For example, it is unclear at which processing stage voice perception difficulties occur, i.e. whether they are rather of apperceptive or associative nature. It is further unclear whether impairments in voice identity processing in ASD are associated with dysfunction of voice-sensitive brain regions, whether they are dissociable from more general impairments in acoustic sound processing, such as music or whether they are associated with deficits in recognising other person related information, such as faces. Within the scope of my dissertation we addressed these open research questions by systematically investigating voice perception in adults with high-functioning ASD and typically developed matched controls (matched pairwise on age, gender, and intellectual abilities). Our results allow, for the first time, a characterisation of the voice perception deficits in ASD. Results from a comprehensive behavioural test battery and two standard functional magnetic resonance imaging (fMRI) experiments that we used to localise voice-sensitive brain regions suggest that in high-functioning ASD: (*i*) difficulties in voice identity and vocal emotion recognition are of perceptual nature, i.e. occur due to difficulties with processing voice acoustic features, such as vocal pitch; (*ii*) difficulties in vocal pitch perception are dissociable from otherwise intact acoustic processing (i.e. hearing ability, musical pitch, musical timbre, and vocal timbre perception); difficulties in voice identity recognition are (*iii*) dissociable from intact speech recognition; (*iv*) associated with impaired face identity recognition; (*v*) on a neuronal level associated with reduced functioning of brain regions associated with voice identity processing and (*vi*) dissociable from a more general neuronal processing deficit of vocal sounds. Our results inform models on human communication and advance our understanding for basic mechanisms which might contribute to core symptoms in ASD, such as difficulties in communication.

Zusammenfassung

Die korrekte Wahrnehmung von Informationen, welche mittels der Stimme transportiert werden, ist eine Grundvoraussetzung erfolgreicher zwischenmenschlicher Kommunikation. Die Stimme einer anderen Person zu hören liefert Informationen darüber wer spricht (Sprechererkennung), was gesagt wird (stimmliche Spracherkennung) und über den emotionalen Zustand einer Person (stimmliche Emotionserkennung). Autismus Spektrum Störungen (ASS) sind mit Einschränkungen in der Stimmenerkennung und der stimmlichen Emotionserkennung assoziiert, während die Wahrnehmung stimmlicher Sprache relativ intakt ist. Die Mechanismen, die diesen Einschränkungen in der Stimmenwahrnehmung zugrunde liegen sind bisher jedoch unbekannt. Es ist beispielsweise unklar, auf welcher Verarbeitungsstufe die Einschränkungen in der Stimmenwahrnehmung entstehen, d.h. ob sie eher apperzeptiver oder assoziativer Natur sind. Des Weiteren ist es unklar, ob Einschränkungen in der Verarbeitung von Sprecherkennung bei ASS mit einer Dysfunktion stimmensensitiver Hirnregionen in Verbindung stehen, ob sie von einer generelleren Einschränkung in der Verarbeitung akustischer Signale, wie der Verarbeitung von Musik, dissoziiert werden können und ob sie mit Einschränkungen in der Erkennung anderer personenrelevanter Information, beispielsweise vom Gesicht, einhergehen. Im Rahmen meiner Dissertation haben wir diese offenen Forschungsfragen adressiert, in dem wir systematisch Stimmenverarbeitung bei Erwachsenen mit hochfunktionalem ASS und typisch entwickelten Kontrollprobanden untersuchten, wobei die Kontrollprobanden in Alter, Geschlecht und intellektuellen Fähigkeiten jeweils vergleichbar mit einem Probanden der ASS Gruppe waren. Unsere Ergebnisse erlauben erstmalig eine Charakterisierung der Einschränkung der Stimmenwahrnehmung bei ASS. Die Ergebnisse einer umfassenden verhaltensbezogenen Testbatterie und zweier Standard funktionaler Magnet Resonanz Tomographie (fMRT) Experimente zur Erfassung stimmensensitiver Hirnregionen deuten darauf hin, dass bei hochfunktionalem ASS: (*i*) Einschränkungen in der Sprechererkennung und der stimmlichen Emotionserkennung perzeptueller Natur sind, d.h. durch Schwierigkeiten mit Verarbeitung stimmlich akustischer Merkmale wie stimmliche Tonhöhe, entstehen; (*ii*) Schwierigkeiten in der Wahrnehmung stimmlicher Tonhöhe von ansonsten intakter akustischer Wahrnehmung (d.h. Hörfähigkeit, Tonhöhenunterscheidung in Musik, stimmlicher und musikalischer Klangfarbenwahrnehmung) unterschieden werden können; Schwierigkeiten in der Sprechererkennung (*iii*) von intakter Spracherkennung unterschieden werden können; (*iv*) mit Schwierigkeiten in der Gesichtererkennung einhergehen; (*v*) auf neuronaler Ebene mit einer reduzierten Aktivität von stimmenerkennungssensitiven Hirnregionen einhergehen und (*vi*) von einem generellen neuronalen Verarbeitungsdefizit vokaler Laute dissoziierbar sind. Unsere Ergebnisse bringen neue Kenntnisse für Modelle zwischenmenschlicher Kommunikation und erhöhen unser Verständnis elementarer

Mechanismen, die den Kernsymptomen in ASS, wie beispielsweise Schwierigkeiten in der Kommunikation, zugrunde liegen könnten.

Table of Contents

1 Introduction

The correct perception of the information carried by the voice is a key requirement to communicate successfully with others (Belin, Fecteau, & Bedard, 2004; Blank, Wieland, & von Kriegstein, 2014; Ellis, Jones, & Mosdell, 1997; von Kriegstein et al., 2008). The accurate interpretation of voice information from conspecifics can be critical for survival. For example, in the animal world, the correct perception of voices can serve to discriminate between 'friend and foe', to keep track of the herd and to be warned in dangerous situations (e.g. Jouventin, Aubin, & Lengagne, 1999; Levey et al., 2009; Magrath, Pitcher, & Gardner, 2009). In human beings, the processing of voice information including the discrimination between different voice identities (e.g. De Casper & Fifer, 1980) or different emotions (e.g. Walker-Andrews & Grolnick, 1983) develops very early in life. Even before birth, fetuses show differential behaviour, i.e. increased heart rate, in response towards their mother's as compared to an unfamiliar voice (Kisilevsky et al., 2003). Moreover, infants use vocal emotion cues in order to guide behaviour in uncertain situations. For example, infants show increased orientation towards their mother in a threatening situation, such as when hearing a fearful voice (Mumme, Fernald, & Herrera, 1996; Vaish & Striano, 2004). In line with the behavioural findings, on the neural level voice-sensitive brain areas are present early within the first year of life (Blasi et al., 2011; Grossmann, Oberecker, Koch, & Friederici, 2010).

Physically, the voice sound is produced by the glottal folds and the vocal tract (Fant, 1960). Air provided by the respiratory system reaches the glottal folds. The vibration of the glottal folds converts this aerodynamic energy into acoustic energy. The resonance properties of the vocal tract then shape this sound into the final voice that we perceive (also see Kreiman, 2011). The interplay of these physical characteristics shapes the individual character of a speaker's voice (e.g. Assaneo et al., 2016; Kreitewolf, Gaudrain, & von Kriegstein, 2014; von Kriegstein, Smith, Patterson, Kiebel, & Griffiths, 2010). As a result, from a person talking we can usually immediately infer characteristics about *who* is speaking (henceforth referred to as *voice identity*). For example, we can infer the approximate age (e.g. Ptacek & Sander, 1966; Waller, Eriksson, & Sorqvist, 2015) or gender (e.g. Meister, Fursen, Streicher, Lang-Roth, & Walger, 2016; Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995) of a person based on his/her voice (for reviews see e.g. Kreiman, 2011; Schweinberger, Kawahara, Simpson, Skuk, & Zaske, 2014). If the person is familiar to us, we might also associate the voice with another individuating feature, such as the name or the face. We usually also can infer the mood of the speaker, for example whether somebody is in a happy or angry mood by *how* something is said (henceforth referred to as *vocal emotion*). As compared to voice identity and vocal emotion, most research focused on another component of voice perception, that is, the investigation of *what* is said, i.e. speech recognition (henceforth referred to as *vocal speech*). Based on a current model on voice processing (Belin et al., 2004) in the following, I will use the term *voice perception* to refer to all these three main components, i.e. voice identity, vocal emotion and vocal speech. The current work mainly focuses on voice identity with excursions to vocal emotion and vocal speech.

Previous studies have shown that all three aspects of voice perception can be impaired and that these impairments can occur because of brain lesions (Assal, Zander, Kremin, & Buttet (1976); Crinion & Price, 2005; Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010; Neuner & Schweinberger, 2000; Peretz et al., 1994; Saygin, Dick, Wilson, Dronkers, & Bates, 2003; Van Lancker & Canter, 1982). Voice perception impairments can also occur in developmental disorders, such as in developmental phonagnosia, a condition which describes an impairment in voice identity recognition in the absence of brain insult (Garrido et al., 2009; Roswandowitz et al., 2014; Xu et al., 2015). Another developmental disorder with impairments in voice perception is autism spectrum disorder (ASD), a condition that is characterised by difficulties in social interactions and communication (American Psychiatric Association [APA], 2013). ASD is associated with impairments in the perception of voice identity (Boucher, Lewis, & Collis, 1998; Klin, 1991; Schelinski, Riedel, & von Kriegstein, 2014) and vocal emotion (Globerson, Amir, Kishon-Rabin, & Golan, 2015; Philip et al., 2010; Rutherford, Baron-Cohen, & Wheelwright, 2002), while the perception of vocal speech is relatively intact, at least when speech is presented under a relatively good signal-to-noise ratio (De Gelder, Vroomen, & van der Heide, 1991; Hillier, Campbell, Keillor, Phillips, & Beversdorf, 2007; Iarocci, Rombough, Yager, Weeks, & Chua, 2010; Irwin, Tornatore, Brancazio, & Whalen, 2011; Schelinski et al., 2014; Smith & Bennetto, 2007; Woynaroski et al., 2013). Although early investigations on voice identity perception were clinical studies on brain-damaged patients (e.g. Assal et al., 1976; Van Lancker & Canter, 1982), current models on voice identity perception are mainly informed by behavioural and neuroimaging studies investigating neurotypically developed individuals (e.g. Belin & Zatorre, 2000; von Kriegstein & Giraud, 2004). Until now, the underlying neural mechanisms of voice identity recognition impairments are relatively unclear (Gervais et al., 2004) and there are only a few studies on developmental deficits on voice identity impairments (Garrido et al., 2009; Roswandowitz et al., 2014). Further, no study precisely and systematically unraveled the patterns and underlying mechanisms of voice identity and vocal emotion recognition impairments in ASD. For example, it is unclear at which processing stage voice perception difficulties are impaired in ASD. That voice recognition can be impaired at several different processing stages has recently been shown by investigating cases of developmental phonagnosia (Roswandowitz et al., 2014). Roswandowitz et al. (2014) showed that voice recognition impairments can be of perceptual nature, i.e. brought about by a failure in the encoding and integration of acoustic voice features on a perceptual level (apperceptive phonagnosia) or arise because of impairments on a cognitive level, i.e. in associating a voice with other person related information, such as a face or a name (associative phonagnosia). In ASD it is further unclear whether impaired voice identity recognition is associated with dysfunction of brain regions that are related to the processing of voice identity or whether it is associated with a more general deficit in the neural processing of voices (including all components of voice perception) as suggested by results of a previous study (Gervais et al., 2004). It is also unclear, at which processing stage vocal emotion recognition difficulties in ASD occur and whether these difficulties might have common mechanisms with impaired voice identity processing.

Targeting these open research questions the current thesis aimed to advance the understanding of voice perception and its impairments by systematically investigating both the behavioural and neural mechanisms of voice processing in ASD and typically developed well matched control participants.

Investigating voice perception in ASD is important because of its social relevance. Impairments in person perception are socially restricting (Fine, 2012; Garcia-Villamisar, Rojahn, Zaja, & Jodra, 2010; Harvey & Penn, 2010; Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008). Investigating the underlying mechanisms of voice perception impairments in ASD would thus provide new insights into the difficulties this population are confronted with in everyday situations. Given the important role of voice perception in communication, voice processing deficits are likely to play a key role in communication difficulties that are a core feature of ASD. A systematic characterisation of voice perception impairments in ASD on a behavioural and neural level not only enhances our understanding of the basic processing mechanisms of communication signals and the functional neuropathology in ASD, but also informs models on human communication (Belin et al., 2004; Blank et al., 2014; Ellis et al., 1997; Young & Bruce, 2011). Potentially, the work will also introduce new approaches for the diagnoses and treatment of ASD symptoms.

The three studies of the present dissertation all investigated voice processing in ASD. I therefore first introduce the relevant concepts (voice model, voice identity, ASD) which served as motivation for my studies. I will discuss the implications of these studies in light of current models on voice perception and research on ASD and provide ideas about future directions which can be derived from the current work.

2 Theoretical Background

2.1 Voice Perception

In the following chapter I will introduce a model on voice perception which served as framework for the current work. Given that the current work mainly focused on voice identity I will then continue to illustrate the current knowledge on the neural mechanisms of voice identity processing.

2.1.1 Models on Voice Perception

Models on voice perception (Belin et al., 2004; Ellis et al., 1997; Figure 1) have been adapted from models originally proposed for face recognition (Bruce & Young, 1986; Calder & Young, 2005). Following these models, the incoming auditory signal is first analysed basically (*general auditory analyses*). After general auditory analyses, the structure of the voice signal is analysed (*voice structural encoding*). Following models on face perception (Bruce & Young, 1986), in the current work the stage of structural encoding includes descriptions of voice features and the configuration of the voice as a whole. Such an acoustic voice feature is for example the fundamental frequency [f0], which is perceived as vocal pitch. These acoustic voice features derive from the physical characteristics of a speaker, i.e. the length of the vocal tracts (vocal tract length [VTL]) and the rate of the vibration of the vocal folds (glottal pulse rate [GPR]) (e.g. Fant, 1960). For example, changes in VTL are perceived as changes in speaker height (Fitch & Giedd, 1999; Smith, Patterson, Turner, Kawahara, & Irino, 2005). Thus, VTL can serve as a cue in speaker identity recognition. Changes in GPR are perceived as changes in vocal pitch (Hanson & Chuang, 1999; Smith & Patterson, 2005), which is an important cue in voice identity (for review see Mathias & von Kriegstein, 2014) and vocal emotion perception (Fairbanks & Pronovost, 1938; Gold et al., 2012; Quam & Swingley, 2012).

The structurally encoded descriptions of the voice signal are further used for processing of the single voice perception components, i.e. voice identity, vocal emotion and vocal speech perception. Thereby, the single components are assumed to be further processed in relatively distinct functional pathways (Belin et al., 2004). For example, there is evidence from functional magnetic resonance imaging [fMRI] studies including neurotypical individuals that although partially overlapping (e.g. Kreitewolf et al., 2014; Lachs & Pisoni, 2004; Perrachione & Wong, 2007; von Kriegstein et al., 2010), vocal speech and voice identity recognition are distinguishable abilities which are processed in relatively independent areas of the human brain (Belin & Zatorre, 2003; Formisano, De Martino, Bonte, & Goebel, 2008; von Kriegstein & Giraud, 2004). Clinical evidence for such dissociations also comes from patient lesion reports and studies with individuals with phonagnosia, describing cases that show a dissociation between voice identity processing and other components of voice perception, i.e. vocal speech recognition (e.g. Assal et al., 1976; Roswandowitz et al., 2014; Van Lancker & Canter, 1982), vocal emotion recognition (Garrido et al., 2009; Hailstone et al., 2010; Roswandowitz et al., 2014), or more general auditory processing, such as intact music perception (Garrido et al., 2009; Roswandowitz et al., 2014), and processing of environmental sounds (Garrido et al., 2009; Neuner & Schweinberger, 2000). Comparably to models on face perception (Bruce & Young, 1986), other person-specific knowledge, such as a person's name is assumed to be processed in an associative modality-free processing component (also see Blank et al., 2014 for current models on person recognition).

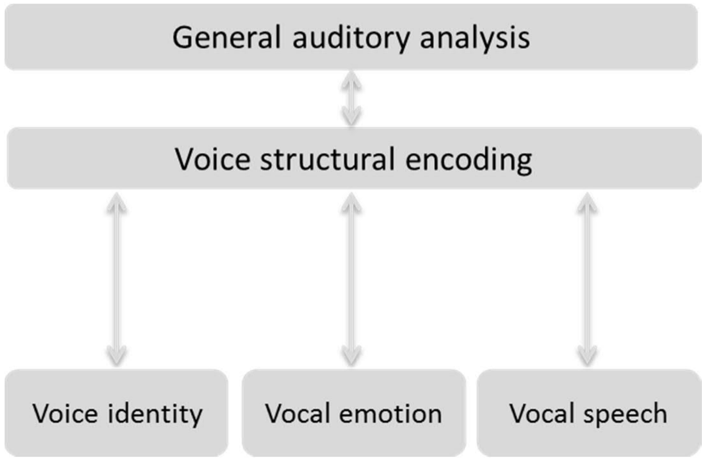

Figure 1

Model of human voice perception (adapted from Belin et al., 2004; Ellis et al., 1997). From top to bottom: After general auditory analyses, the voice signal is encoded structurally, including a provision of descriptions of voice acoustic features, such as the fundamental frequency which is perceived as vocal pitch. The model assumes that voice identity, vocal emotion, and vocal speech are further processed in partially independent functional pathways.

Because the main focus of the current thesis is identity processing, I mainly concentrate on voice identity processing in the following chapter. For completeness, I would like to point to reviews providing comprehensive information on vocal emotion (e.g. Frühholz & Grandjean, 2013; Schirmer & Kotz, 2006) and vocal speech recognition (e.g. Hickok & Poeppel, 2007; Price, 2012).

2.1.2 Neural Mechanisms of Voice Identity Perception

First empirical evidence for neural mechanisms of voice identity perception came from clinical studies investigating patients with brain lesions. These early studies provided behavioural evidence that impaired voice identity processing is associated with brain lesions (e.g. Assal et al., 1976; Van Lancker & Kreiman, 1987; Van Lancker & Canter, 1982) including bilateral temporal and right parietal lobe lesions (e.g. Van Lancker, Cummings, Kreiman, & Dobkin, 1988; Van Lancker, Kreiman, & Cummings, 1989). Later studies with lesion patients highlighted a predominant role of the right hemisphere in voice identity processing (Lang, Kneidl, Hielscher-Fastabend, & Heckmann, 2009; Neuner & Schweinberger, 2000) and the role of the anterior part of the right temporal lobe in multimodal person recognition including voice-, face-, and name recognition (Hailstone et al., 2010; Hailstone et al., 2011).

First fMRI studies with typically developed individuals identified the right superior temporal sulcus and gyrus [STS/STG] as the main voice-sensitive brain region (e.g. Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; von Kriegstein & Giraud, 2004). These fMRI studies on voice perception used two different approaches which became standard to localise voice-sensitive brain regions: In one approach, participants passively listen to blocks of vocal (e.g. coughing, speech, emotion) and non-vocal sounds (e.g. car or bell sounds) (Belin et al., 2002; Belin et al., 2000). Typically developed individuals usually show higher responses to vocal as compared to non-vocal sounds in voice-sensitive regions that are located in the bilateral STS/STG with right hemispheric dominance. In this approach, blocks of vocal sounds include all three main components on voice perception (Figure 1), i.e. blocks can include different sentences spoken by different speakers (voice identity), spoken with different emotional expressions (vocal emotion, e.g. laughing) and spoken in different languages (vocal speech). The other approach to localise voice-sensitive brain regions more specifically targets voice identity recognition. This approach identifies voice-sensitive brain regions along the right STS/STG that preferably respond when participants perform a voice identity recognition task as compared to a vocal speech recognition task on the same stimulus material (von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004).

A range of further studies confirmed the role of the STS/STG as main voice-sensitive brain region and refined its role in typically developed individuals using fMRI (e.g. Andics et al., 2010; Belin & Zatorre, 2003; Bonte, Hausfeld, Scharke, Valente, & Formisano, 2014; Formisano et al., 2008) and other functional neuroimaging techniques such as Electroencephalography [EEG] (e.g. Charest et al., 2009) and Magnetoencephalography [MEG] (e.g. Schall, Kiebel, Maess, & von Kriegstein, 2015). Further, a study using repetitive transcranial magnetic stimulation [rTMS] revealed a causal role of the right STS/STG in voice perception (Bestelmeyer, Belin, & Grosbras, 2011). This study showed that TMS of the right STS/STG, i.e. disrupting its activity, impaired discrimination between vocal and non-vocal sounds. Recent functional studies showed that in humans these voice-sensitive brain regions develop early in infancy (Blasi et al., 2011; Grossmann et al., 2010) and refine until adulthood (Bonte et al., 2013; Bonte, Ley, Scharke, & Formisano, 2016). The existence of voice-sensitive brain regions has also been shown in non-human primates (e.g. Petkov et al., 2008) and other animals like dogs (Andics, Gacsi, Farago, Kis, & Miklosi, 2014; for review see Perrodin, Kayser, Abel, Logothetis, & Petkov, 2015).

Along the right voice-sensitive STS/STG previous research identified functionally different sub-regions. The posterior STS/STG has been associated with the processing of complex spectrotemporal voice features (e.g. Andics et al., 2010; Blank et al., 2014; von Kriegstein & Giraud, 2004; von Kriegstein et al., 2010; Warren, Scott, Price, & Griffiths, 2006), while the anterior part has been more closely related to voice identity processing (e.g. Andics et al., 2010; Belin & Zatorre, 2003; Schall et al., 2015; von Kriegstein & Giraud, 2004).

Besides the STS/STG, other brain regions located in the auditory cortex, such as Heschl's gyrus (Bonte et al., 2014; Formisano et al., 2008) and the planum temporale (von Kriegstein & Giraud, 2006; Warren et al., 2006) have been identified as part of the so called ‘core’ voice-system in voice processing. Voice-identity processing also leads to responses in supramodal brain regions which belong to the so called ‘extended’ voice system, such as the precuneus/posterior cingulate, the temporal pole, or the amygdala, inferior frontal gyrus, and regions of other sensory modalities, such as the fusiform face area (Andics et al., 2010; Latinus, Crabbe, & Belin, 2011; Schall, Kiebel, Maess, & von Kriegstein, 2013; Shah et al., 2001; von Kriegstein & Giraud, 2006; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; for review see Blank et al., 2014). These regions are functionally or/and structurally connected to regions of the core-voice system (Blank, Anwander, & von Kriegstein, 2011; Blank, Kiebel, & von Kriegstein, 2015; von Kriegstein & Giraud, 2006; von Kriegstein et al., 2005).

2.2 Autism Spectrum Disorder

The term autism originates from the Greek word *αὐτός* (autos), meaning *self* and the suffix *ισμός* (ismós), meaning *state* or *condition* and was originally used by the Swiss psychiatrist Eugen Bleuer to describe the tendency in schizophrenic patients to withdraw into inner life with an active turning-away from the external world (Bleuler, 1911). In the 1940s, the term autism was used in two pioneering works to describe children who showed marked qualitative abnormalities in reciprocal social interactions and communication, and restricted, stereotyped, repetitive repertoire of interests and activities that were present in early childhood (Kanner, 1943; Asperger, 1944). The diagnostic classification of ASD, i.e. Kanner syndrome/ childhood autism and Asperger syndrome, was based on these descriptions and the described symptoms are still core diagnostic features of ASD in current standard classification systems (APA, 2013; World Health Organisation [WHO], 2004). In the most current version of the Diagnostic and Statistical Manual of Mental Disorders [DSM-5] (APA, 2013), the *spectrum* concept of autism was introduced, encompassing the previous diagnoses including Kanner and Asperger syndrome in favor of a dimensional approach which classifies individuals on the autism spectrum based on the severity in the symptom domains. According to the level of 'cognitive functioning' or intellectual abilities, ASD is also often labelled as being 'high-functioning' or 'low-functioning'. The term 'high-functioning autism' usually describes individuals with an intelligence quotient [IQ] above 85, whereas the term 'low-functioning autism' usually describes individuals with an IQ below 70 (Baron-Cohen, 2008).

With prevalence estimates around 1% (for review see Fombonne, 2009) or even higher (Kim et al., 2011; Weintraub, 2011), the occurrence of ASD in the general population is relatively high and has a significant impact on life of individuals, families, and various sectors of society (Baxter et al., 2015; Knapp, Romeo, & Beecham, 2009). Typically, ASD is more often diagnosed in males than females (e.g. Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Fombonne, 2009; Kim et al., 2011). There is good evidence that genetic factors contribute to ASD (e.g. Lichtenstein, Carlstrom, Rastam, Gillberg, & Anckarsater, 2010). However, the full etiology is currently unclear and other factors, such as neurobiological and environmental factors are discussed as potential significant contributors to ASD (for reviews see e.g. Barak & Feng, 2016; Hertz-Picciotto et al., 2006; Tick, Bolton, Happe, Rutter, & Rijsdijk, 2016). For example, ASD has been associated with abnormal brain structure, including abnormal local anatomical structures and altered connectivity between brain regions (for review see e.g. Ismail et al., 2016; Just, Keller, Malave, Kana, & Varma, 2012; Maximo, Cadena, & Kana, 2014; Nickl-Jockschat et al., 2012). Further there is a large number of studies highlighting differences in brain functions that are related to social cognition, i.e. differences in the social brain network in ASD including the STS/STG region (for reviews in ASD see e.g. Dziobek & Kohne, 2011; Pelphrey, Shultz, Hudac, & Vander Wyk, 2011). In a broad sense, social cognition refers to processing that is elicited by or directed towards other people and includes cognitive processing (e.g., perception, reasoning, memory, attention, motivation, and decision-making) that underlies a social ability or social behavior (for review see Kennedy & Adolphs, 2012). Voice, face, and emotion processing are integral parts of this social cognition network (Happe & Frith, 2014). Within the scope of social cognition, prominent theories focus in higher cognitive functions, such as accentuating that individuals with ASD have impairments in making inferences about one's own and other people's mental states ('theory of mind'; Baron-Cohen, Leslie, & Frith, 1985) or show diminished motivation towards social stimuli (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). Other prominent theories on ASD, such as the theory of weak central coherence (Frith, 1989), or the theory of executive dysfunction (Ozonoff, Pennington, & Rogers, 1991), emphasise more general different cognitive processing styles in ASD. The theory of weak central coherence describes a reduced tendency in ASD to integrate local information into a coherent or 'global' and meaningful whole (Frith, 1989; Frith & Happe, 1994; Happe & Frith, 2006). Of particular importance within the scope of the current work are further approaches for ASD which emphasise the importance of sensory processing in characterising and understanding ASD (for review see Baum, Stevenson, & Wallace, 2015; Dakin & Frith, 2005; Kern et al., 2007; Pellicano & Burr, 2012). Although sensory dysfunctions are now also integral parts in the DSM-5, sensory contributions to ASD symptomatology and impairments in higher social cognition have been poorly recognised and poorly characterised (for review see Baum et al., 2015).

2.3 Voice Perception in Autism Spectrum Disorder

There is evidence that the processing of single components of voice perception is impaired in ASD (for review see O'Connor, 2012). For example, there is evidence that individuals with ASD are impaired in recognising emotions from voices (Globerson et al., 2015; Golan, Baron-Cohen, Hill, & Rutherford, 2007; Philip et al., 2010; Rosenblau, Kliemann, Dziobek, & Heekeren, 2017; Rutherford et al., 2002; for review see Lartseva, Dijkstra, & Buitelaar, 2015). However, there are only a few studies on voice identity recognition in ASD (Boucher et al., 1998; Boucher, Lewis, & Collis, 2000; Lin et al., 2015; Schelinski et al., 2014). Early evidence for atypical voice identity processing came from a study which showed that children with ASD lack preference for their mother's voice (Klin, 1991). Boucher and colleagues (1998) showed that the ability to discriminate personally familiar voices (i.e. staff from the children's own school) from unfamiliar voices (i.e. unknown staff from a different school) was impaired in children with ASD, in comparison to children without ASD but with language delay. The voice recognition impairments in ASD children could not be replicated in a further study which employed a simplified task and stimuli which were easier to discriminate. It was speculated that these difficulties may be particularly evident when task demands are high and voices are difficult to discriminate (see Boucher et al., 2000). In a previous study, we showed that adults with high-functioning ASD are impaired in recognising voice identity from speakers who have been learned together with a name or together with a symbol of their occupation (Schelinski et al., 2014; but also see Lin et al., 2015, who found intact voice-name learning in adults with ASD). Other studies provided evidence that processing of further voice features which can be used to infer identity, such as recognising gender or age by voice, is intact in ASD (Clopper, Rohrbeck, & Wagner, 2013; Groen et al., 2008; Lin et al., 2015). Further, in a previous study we found evidence that voice recognition impairments in adults with high-functioning ASD are dissociable from vocal speech recognition performances, which were found to be intact in that population (Schelinski et al., 2014). That vocal speech recognition is intact in ASD, at least when speech is presented under a relatively good signal-to-noise ratio, is in line with other previous studies that reported similar speech recognition performances between ASD participants and controls (De Gelder et al., 1991; Hillier et al., 2007; Iarocci et al., 2010; Irwin et al., 2011; Smith & Bennetto, 2007; Woynaroski et al., 2013).

On the neural level, the STS/STG region, conceptualised as part of the social brain network, has also been proposed to play a critical role in explaining ASD symptomatology (for reviews see e.g. Dakin & Frith, 2005; Redcay, 2008; Yang, Rosenblau, Keifer, & Pelphrey, 2015; Zilbovicius et al., 2006). One often cited study showed that adults with ASD showed no increased brain responses in the voice-sensitive STS/STG region when passively listening to vocal as compared to non-vocal sounds (Gervais et al., 2004). More recently a study using resting state fMRI showed that this voice-sensitive brain region exhibited weaker connectivity with brain structures that are involved in reward and emotion processing in ASD as compared to control participants (Abrams et al., 2013). The results by Gervais et al. (2004) suggest more general reduced cortical processing of the voice in ASD. However, these results are based on a small number of participants (n_{ASD} = 5). Further, as blocks of vocal sounds include stimuli conveying voice identity (e.g. sentences spoken by different speakers) but also vocal emotion (e.g. laughs, cries) and vocal speech (e.g. sentences spoken in different languages), it remains speculative whether the reduced brain response in voice-sensitive regions reflected a selective deficit in voice identity processing or a general lack of processing of vocal sounds including identity, emotion and speech information. Thus, the neural basis of selective voice identity processing in ASD is still subject to debate.

Summing up the current knowledge on voice identity processing in ASD, the underlying mechanisms of impairments in voice identity recognition are unclear. For example, it is unclear at which processing stage impairments occur: Studies on phonagnosia suggest that voice identity processing deficits can occur at different stages of processing, according to which phonagnosia can be classified as being of an apperceptive or associative nature (Hailstone et al., 2010; Hailstone et al., 2011; Roswandowitz et al., 2014), a classic distinction in agnosia research (Buchtel & Stewart, 1989; Lissauer, 1890). In apperceptive phonagnosia, voice identity processing deficits arise because of perceptual processing deficits, i.e. deficits in perceiving and integrating acoustic voice features (Roswandowitz et al., 2014). In associative phonagnosia voice recognition impairments arise because of impairments in associating a voice with other person identity information such as the face or name (Hailstone et al., 2010; Roswandowitz et al., 2014). Another interesting finding in a recent study by Roswandowitz et al. (2014) was that developmental phonagnosia is associated with a deficit in vocal pitch discrimination, whereas vocal timbre and non-vocal pitch perception were intact. This might suggest that poor vocal pitch perception exacerbates impaired voice identity processing (Roswandowitz et al., 2014) and supports the relevance vocal pitch as a key acoustical voice property individuating the human voice as shown in previous studies in typically developed individuals (e.g. Abberton & Fourcin, 1978; Gaudrain, Li, Ban, & Patterson, 2009; Kreitewolf et al., 2014). Impaired vocal pitch perception has also been associated with impairments in vocal emotion recognition in individuals with schizophrenia (Gold et al., 2012). With regard to the analysis of acoustical voice features in ASD the majority of studies focused on pitch processing and neglected a potential relevance of vocal timbre processing during voice-identity processing. Those studies indicated normal or even enhanced processing of vocal and non-vocal pitch in ASD (e.g. Bonnel et al., 2010; Globerson et al., 2015; Jarvinen-Pasley & Heaton, 2007; for review see O'Connor, 2012). Only a few studies directly compared the ability to process vocal and non-vocal pitch within the same sample of participants: One study provided evidence that in ASD non-vocal pitch perception (i.e., judging whether a single tone changed in pitch in pairs of melodies) is intact and dissociable from impaired vocal pitch perception (i.e., judging differences in intonation) (Jiang, Liu, Wan, & Jiang, 2015). This finding contrasted another study reporting that individuals with ASD showed intact non-vocal pitch perception (i.e., judging sameness of pairs of tones) and were even more sensitive in vocal pitch perception (i.e., judging sameness of pairs of words) as compared to typically developed individuals (Jarvinen-Pasley & Heaton, 2007). There are several factors that could explain the discrepancy between the findings, such as differences in the sample characteristics (e.g., differences in age or type of ASD diagnosis) and task design (e.g., differences in task difficulty and instruction). To our knowledge, only one study tested timbre perception in ASD (Bonnel et al., 2010). This study found intact timbre processing in ASD for vocal and non-vocal sounds.

The prominent weak central coherence view on ASD is mainly based on observations from the visual domain (Frith & Happe, 1994; Happe & Frith, 2006). If and how this theoretical framework can be translated to the auditory modality is currently under discussion (for review see Haesen, Boets, & Wagemans, 2011). Recently, it has been suggested that the importance of altered sensory processing in ASD should be highlighted in order to better understand ASD symptomatology. In this view, the impact of altered lower-level sensory processing in understanding differences in higher-order processing is highlighted (Baum et al., 2015). In the view of voice perception, differences in perceptual processing, such as of vocal pitch, might explain higher cognitive impairments in voice perception, such as voice identity or vocal emotion recognition. For example, in a previous study perceptual processing of non-vocal pitch sounds was associated with impairments in vocal emotion processing (Globerson et al., 2015). However, there is no systematic investigation on the direct relation between perceptual and higher cognitive vocal sound processing, i.e. whether processing of acoustic voice features relates to impaired higher cognitive voice processing. Demonstrating an association between perceptual processing and higher cognitive processing difficulties in ASD might highlight the need to consider perceptual processing difficulties as a significant contributor to at least some of the social cognitive impairments that are integral part of prominent theories in ASD.

2.4 Rationale of the Empirical Studies

The aim of the current dissertation was to inform models on voice perception by investigations on ASD. The studies also contribute to a better understanding of communication mechanisms in ASD. In the following paragraph I introduce the research questions that are addressed in the three empirical studies of the current thesis.

In the current studies, we systematically investigated voice perception with a focus on voice identity recognition in adults with high-functioning ASD and typically developed pairwise matched controls. The first research aim of the current thesis was to characterise the voice identity processing deficit in ASD and how it relates to vocal emotion and vocal speech processing. **Study 1** and parts of **study 2** focused on the behavioural characterisation of voice identity processing in ASD. Within the scope of this characterisation we addressed several research questions which arose from the current knowledge on voice processing (see chapter 2.1): *(i)* At which processing stage do voice identity processing impairments in ASD occur, i.e. are impairments rather of an apperceptive or associative nature? Are the impairments in voice identity processing in ASD *(ii)* dissociable from more general acoustic processing of sounds, such as music sounds?; *(iii)* dissociable from intact vocal speech recognition, replicating our previous findings (Schelinski et al., 2014)? *(iv)* dissociable from deficits in recognising other person related information, such as faces?

Study 2 aimed to answer the research question whether voice-sensitive STS/STG regions are involved in voice identity processing deficits in ASD. The STS/STG is a well known structure involved in voice perception (see chapter 2.1.2). However, the functional role of this region in voice identity processing impairments is unclear. As part of the 'social brain' the STS/STG is also known as a key structure in explaining ASD symptomatology (see chapters 2.2 and 2.3). Results of a previous study suggest that voice-sensitive STS/STG regions are absent in ASD (Gervais et al., 2004). However, whether this finding can be replicated in a bigger and well matched sample and whether it can explain behavioural impairments in voice identity recognition in ASD is unclear. **Study 2** addressed these questions by investigating, for the first time, whether behavioural voice identity recognition impairments are associated with dysfunction in voice-sensitive STS/STG regions in a well characterised ASD sample and well matched typically developed control participants.

Study 3 addressed the research question whether impairments in perceiving acoustical voice features are related to impaired vocal emotion perception. Besides impaired voice identity perception, ASD is associated with deficient vocal emotion recognition (see chapter 2.3). A key acoustic voice feature involved in voice identity and vocal emotion recognition is the fundamental frequency which is perceived as vocal pitch (see chapter 2.1.1). We previously suggested that abilities in vocal pitch processing impact on voice identity recognition performance (Roswandowitz et al., 2014). However, the role of vocal pitch processing in vocal emotion recognition impairments is unclear. **Study 1** and **study 3** addressed the role of processing acoustical voice features, i.e. vocal pitch, during voice identity (study 1) and vocal emotion recognition (study 3) in ASD.

3 Empirical Studies

3.1 Summaries

3.1.1 Study 1: *Voice Identity Processing in Autism Spectrum Disorder*

Recognising the identity of others is a key requirement for successful human communication (Belin et al., 2004; Blank et al., 2014; Bruce & Young, 1986; Ellis et al., 1997). Previous research suggests that voice identity processing is impaired in ASD (Boucher et al., 1998; Klin, 1991; Schelinski et al., 2014). However, the characteristics and underlying perceptual mechanisms of this voice processing deficit are unknown. In study 1 we systematically investigated voice processing in ASD. Sixteen adults with high-functioning ASD and sixteen typically developed controls (matched pairwise on age, gender, and IQ) participated in a comprehensive behavioural test battery which included tests on *(i)* unfamiliar voice discrimination, newly learned voice recognition, and famous voice recognition; *(ii)* acoustical voice processing abilities that are associated with voice identity processing (i.e. tests on vocal pitch and timbre discrimination); *(iii)* control tasks (i.e. tests on hearing abilities, musical pitch and timbre perception, and face identity recognition). The results showed that the ASD group was particularly impaired in discriminating, learning, and recognising unfamiliar voices, while performance in recognising famous voices was not significantly different from controls. The behavioural impairment with unfamiliar voices correlated with an impairment in recognising unfamiliar faces. Tests on acoustic processing abilities showed that the ASD group had a specific deficit in vocal pitch discrimination that was dissociable from otherwise intact acoustic processing (i.e. hearing ability, musical pitch, musical timbre, and vocal timbre perception tests). These results allow, for the first time, a characterisation of the voice recognition deficit in ASD: The findings indicate that in high-functioning ASD, *(i)* the difficulty to recognise voices is particularly pronounced for learning and recognising novel voices; *(ii)* the ability to recognise voices is associated with impaired processing of acoustic aspects of voices on the perceptual level, that is, impairments with integrating the acoustic characteristics of the voice into a coherent percept; *(iii)* standard voice perception models need to be amended to incorporate distinctions between apperceptive and associative processes.

3.1.2 Study 2: *Temporal Voice Areas Exist in Autism Spectrum Disorder but have Selective Dysfunction for Voice Identity Recognition*

Brain areas that respond selectively to human voice sounds have been identified along the STS/STG (Belin et al., 2002; Belin et al., 2000) and regions along the right STS/STG are thought to serve different functions in voice identity recognition (Andics et al., 2010; Belin & Zatorre, 2003; Schall et al., 2015; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004). Currently it is unclear whether dysfunction of these voice-sensitive regions can explain voice identity recognition impairments. Study 2 investigated whether dysfunction of voice-sensitive regions can explain the voice identity processing deficits found in ASD (Schelinski, Roswandowitz, & von Kriegstein, 2017; study 1). A previous study (Gervais et al., 2004) made the often cited claim that individuals with ASD generally fail to activate voice-sensitive regions in response to vocal sounds. Study 2 systematically addressed the neural mechanisms of voice processing in ASD using two established approaches to localise voice-sensitive brain regions in larger and well matched samples. Sixteen adults with high-functioning ASD and sixteen typically developed controls (matched pairwise on age, gender, and IQ) participated in two fMRI experiments. In the vocal sound experiment, participants passively listened to blocks of vocal (speech and non-speech) and non-vocal sounds (e.g. musical instruments, nature, animals). In the voice identity recognition experiment, participants listened to sentences spoken by three male speakers and performed voice identity recognition and vocal speech recognition tasks on the same stimuli. Here, participants decided whether a speaker matched the identity of a target speaker (voice identity task) or whether the content of a sentence matched the content of a target sentence (speech task). In the vocal sound experiment there was greater voice-sensitive blood-oxygenation-level-dependent [BOLD] response along the bilateral STS/STG when participants listened to vocal as compared to non-vocal sounds in the ASD as well as in the control group. In contrast, in the voice identity recognition experiment there was enhanced BOLD response in the voice-sensitive right posterior STS/STG when recognising voice identity as compared to vocal speech in controls compared to the ASD group. Responses in the right anterior STS/STG correlated positively with voice identity recognition behavioural performance in controls but not in the ASD group. Replicating previous results (Schelinski et al., 2014), on the behavioural level, the ASD group performed worse than controls in the voice identity recognition task whereas both groups performed equally well in the vocal speech recognition task. Contrasting previous results (Gervais et al., 2004) the findings clearly show that individuals with high-functioning ASD exhibit BOLD responses in voice-sensitive regions in response to vocal sounds, thus in disagreement with the idea that ASD is characterised by a general lack of response in voice-sensitive areas. However, the voice-sensitive regions showed a partial dysfunction that can explain the specific recognition impairments for unfamiliar voices in ASD. Altered functioning of voice identity processing, that is distinguishable from vocal speech processing, contributes to better understanding the neural mechanisms that might at least partly underlie communication deficits in ASD. In addition the results provide insights into the behavioural relevance of voice-sensitive regions in the STS/STG: We provide evidence that voice-sensitive regions in the brain play a critical role in the recognition of voice identity, beyond their role in discriminating voices from non-vocal sounds (Gervais et al., 2004). Furthermore the results of study 2 support and extend the findings of study 1, suggesting that it is the posterior temporal lobe voice region that integrates the acoustic characteristics of the voice into a coherent percept – a process that is an integral part of the most influential voice perception model (Belin et al., 2004).

3.1.3 Study 3: *Vocal Pitch Perception Abilities are Associated with Emotion Recognition Abilities in Controls but not in Autism Spectrum Disorder*

The ability to recognise emotions expressed by the voice is another component critical for successful communication (Belin et al., 2004). Besides impaired voice identity perception, ASD is associated with deficient vocal emotion recognition (Globerson et al., 2015; Philip et al., 2010; Rutherford et al., 2002). Previous study results suggest that abilities in vocal pitch processing impact on voice identity recognition impairments (Roswandowitz et al., 2014). The fundamental frequency which is perceived as vocal pitch is also a key acoustic feature in vocal emotion recognition (Fairbanks & Pronovost, 1938; Gold et al., 2012; Quam & Swingley, 2012). However, whether impaired vocal pitch processing also impacts vocal emotion recognition impairments is currently unclear. Study 3 aimed to investigate whether impaired perception of vocal pitch is related to impaired vocal emotion processing in ASD. Individuals with high-functioning ASD (*n* = 16) with known impairments in vocal pitch perception and intact non-vocal pitch perception abilities (study 1) and pairwise matched typically developed individuals (*n* = 16) performed a test on vocal emotion recognition. The ASD group performed worse in vocal emotion recognition as compared to the control group. Within the control group performance in the vocal emotion recognition task positively correlated with performance in vocal pitch, but not non-vocal pitch perception. This indicated that in controls better vocal emotion recognition abilities were associated with better vocal pitch perception. Additionally, within the control group higher extends of traits associated with the autism spectrum were associated with worse performance in vocal emotion recognition. There were no correlations between vocal emotion recognition abilities and the ability to perceive vocal and non-vocal acoustic parameters within the ASD group. The results of study 3 highlight the relevance of impaired vocal emotion recognition as a characteristic of high-functioning ASD. The findings suggest that perceptual impairments, i.e. a deficit in vocal pitch perception might contribute to impairments in vocal emotion recognition in ASD. Further, we speculate that impairments in vocal emotion and voice identity recognition in high-functioning ASD might have an at least partly common origin, underpinned by dysfunction of the posterior STS/STG.

3.2 Manuscripts

3.2.1 Publication Study 1

Voice Identity Processing in Autism Spectrum Disorder

Stefanie Schelinski[1], Claudia Roswandowitz[1,2] & Katharina von Kriegstein[1,3]

[1] *Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, Leipzig, Germany*

[2] *International Max Planck Research School on Neuroscience of Communication, Stephanstraße 1a, Leipzig, Germany*

[3] *Humboldt University of Berlin, Institute of Psychology, Rudower Chaussee 18, Berlin, Germany*

Published in

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research, 10*(1), 155-168.

10.1002/aur.1639

3.2.2 Publication Study 2

Temporal Voice Areas Exist in Autism Spectrum Disorder but are Dysfunctional for Voice Identity Recognition

Stefanie Schelinski[1], Kamila Borowiak[1,2] & Katharina von Kriegstein[1,3]

[1] *Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, Leipzig, Germany*

[2] *Berlin School of Mind and Brain, Humboldt University of Berlin, Luisenstraße 56, Berlin, Germany*

[3] *Humboldt University of Berlin, Institute of Psychology, Rudower Chaussee 18, Berlin, Germany*

Published in

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.

doi: 10.1093/scan/nsw089

3.2.3 Manuscript Study 3

Vocal Emotion Difficulties in Autism Spectrum Disorder are Associated with Impaired Vocal Pitch Processing

Stefanie Schelinski[1] & Katharina von Kriegstein[1,2]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, Leipzig, Germany

[2] Technische Universität Dresden, Zellescher Weg 17, Dresden, Germany

Manuscript submitted for publication

Manuscript available from first author

4 General Discussion

Our studies provide new knowledge on voice processing in high-functioning ASD and enhance our understanding of basic communication mechanisms within this condition. In summary, our work demonstrates that adults with high-functioning ASD *(i)* are impaired in processing unfamiliar voice identities but not in associating person related information to familiar voices, indicating that the voice identity recognition deficit is of perceptual nature; *(ii)* show a voice selective deficit in pitch processing, one main voice acoustic feature that determines voice perception; *(iii)* have a dysfunction in selective parts of the temporal lobe known to be sensitive to voice identity. The results show that these impairments in voice identity processing are *(iv)* dissociable from more general intact basic acoustic processing (i.e. hearing abilities, music perception); *(v)* dissociable from intact performances in vocal speech recognition; *(vi)* associated with impaired face identity recognition and *(vii)* on a neural level dissociable from a more general processing deficit of vocal sounds. Further the results show that *(viii)* the ASD group had impairments in vocal emotion recognition that might at least partly be explainable by impaired vocal pitch perception. The behavioural and neuroimaging results converge to support the view that a specific deficit in integrating the acoustic characteristics of the voice into a coherent percept is related to high-order impairments in voice perception i.e. voice identity and vocal emotion recognition in ASD.

4.1 Implications for Models on Voice Perception

With regard to current models on voice perception, the current work lends important evidence to support some core claims and provides suggestions for modification. The current thesis shows for the first time that a voice identity processing deficit is associated with a dysfunction of voice-sensitive right STS/STG regions. In line with previous studies, the behavioural and neural findings support the role of the posterior part of the STS/STG in processing complex spectrotemporal features (e.g. Andics et al., 2010; Belin & Zatorre, 2003; von Kriegstein & Giraud, 2004; von Kriegstein et al., 2010; Warren et al., 2006). The results from study 2 additionally support the role of the anterior part of the STS/STG in voice identity recognition (e.g. Andics et al., 2010; Belin & Zatorre, 2003; Schall et al., 2015; von Kriegstein & Giraud, 2004).

The results of the current work further support a central assumption of models on human auditory communication (Belin et al., 2004; von Kriegstein et al., 2008): Although partially overlapping (e.g. Kreitewolf et al., 2014; Lachs & Pisoni, 2004; Perrachione & Wong, 2007; von Kriegstein et al., 2010), voice identity and vocal speech recognition are distinguishable abilities which are processed relatively independently in the human brain (Belin & Zatorre, 2003; Formisano et al., 2008; von Kriegstein & Giraud, 2004).

The results of the present work suggest that standard voice recognition models (Belin et al., 2004; Blank et al., 2014) might need to be amended to incorporate distinctions between unfamiliar and familiar voice processing at the perceptual level. Our results suggest that people with high-functioning ASD might have difficulties in analysing and integrating the complex acoustic voice features and that this might be particularly important for learning novel voices, whereas recognising familiar voices (i.e. famous) voices might rely on a partially different mechanisms. These results are in line with previous findings that identity processing i.e. discriminating different voices (perceptual voice processing) and associating semantic information to a familiar voice (associative voice processing) can be impaired independently (e.g. Garrido et al., 2009; Hailstone et al., 2010; Van Lancker & Kreiman, 1987; Van Lancker et al., 1988). Thus, in line with a recent study on developmental phonagnosia our results support the finding that voice identity processing can be impaired on different processing stages (Roswandowitz et al., 2014; Figure 2): On the one hand, voice recognition impairments can be of perceptual nature, i.e. brought about by a failure in the encoding and integration of acoustical voice features in an early processing stage (apperceptive phonagnosia) or arise because of impairments on a cognitive level, i.e. in associating a voice with other person related information, such as a face or a name (i.e. semantic association; associative phonagnosia). These two stages correspond to the classic distinction into apperceptive and associative agnosias (Buchtel & Stewart, 1989; Lissauer, 1890). In the following chapter, I will introduce a working model on voice perception in ASD which incoorperates these findings based on a new adapted model on voice perception (Roswandowitz, Maguinness, & von Kriegstein, under review).

With regard to mechanisms which underlie higher cognitive voice perception, the results suggest that impaired voice identity and vocal emotion recognition is related to difficulties in perceiving vocal pitch and support previous findings that indicate that vocal pitch information is essential for differentiating and recognising vocal emotion (e.g. Fairbanks & Pronovost, 1938; Gold et al., 2012; Quam & Swingley, 2012; Scherer, Banse, Wallbott, & Goldbeck, 1991). Further, the current work suggests partly common mechanisms for voice identity and vocal emotion recognition which might appear at the perceptual level, i.e. when the voice signal is encoded structurally (Figure 2), a mechanism that was previously proposed for face processing (Calder & Young, 2005). Previous models on face perception suggest functionally distinct pathways for the processing of face identity and face emotion that occurs already at early processing stages, i.e. the structural encoding of the face (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000). Amending these models, Calder and Young (2005) suggested, that although overlapping there is some separation between the coding of face identity and face emotion and that the dominant view on distinctive processing is not strongly supported. The authors suggest a framework in which the perceptual representations of both face identity and face emotion are coded by a partly common representational system (Calder & Young, 2005). Evidence for this view comes for example from clinical studies, such as from cases of prosopagnosia, in which besides face identity impairments, face emotion recognition is also often affected (for review see Calder & Young, 2005). Showing that in high-functioning ASD, both, voice identity and vocal emotion recognition was affected, we provide clinical evidence that a similar mechanism as suggested for face perception might also hold for voice perception, i.e. that voice identity and vocal emotion processing are not processed entirely distinctly but share at least some common processing mechanisms at an early processing stage when the voice is encoded structurally.

4.2 Implications for Autism Spectrum Disorder

The findings of this current thesis also advance our understanding of how people with ASD communicate by characterising the behavioural and neural mechanisms of voice perception.

The current work supports and advances previous studies which showed impairments in voice identity (Boucher et al., 1998; Klin, 1991; Schelinski et al., 2014) and vocal emotion perception in ASD (Globerson et al., 2015; Golan et al., 2007; Philip et al., 2010; Rutherford et al., 2002). The apperceptive nature of the voice recognition impairment indicated that these difficulties might arise because of problems with the perceptual analysis of the vocal signal, such as difficulties in processing vocal pitch. Contrasting a previous prominent view, the difficulties in voice identity perception in ASD as identified in the current work cannot be explained by a general deficit in voice processing (Gervais et al., 2004), nor by a general basic auditory processing deficit.

We provide a working model on voice perception in ASD (Figure 2) which is based on a newly introduced model on voice perception (Roswandowitz et al., under review). This new model on voice perception is an adapted and extended version of the model on voice perception which was the framework of the current work (Belin et al., 2004; see 2.1.1), incorporating recent findings on developmental phonagnosia (Roswandowitz et al., 2014; Roswandowitz, Schelinski, & von Kriegstein, 2017), i.e. the multistage nature of voice identity processing. The model assumes that voice identity recognition impairments can emerge at different processing stages, i.e. can be of an apperceptive or associative nature. In apperceptive phonagnosia voice identity processing impairments may emerge due to dysfunction at an early perceptual stage when the voice is encoded structurally. Associative phonagnosia is assumed to be due to dysfunction at a later processing stage, i.e. when a familiar voice is linked with stored semantic information (semantic association) (Roswandowitz et al., 2017). Results from the current work suggest that difficulties in voice identity and vocal emotion in ASD may also emerge due to dysfunction at an early processing stage implicating that they are apperceptive in nature. Poor structural analysis of the voice individuating properties, i.e. vocal pitch may result in a weak representation of the voice percept which might lead to voice recognition difficulties at later processing stages. This might be especially the case for unfamiliar voices, which require more structural analyses than familiar voices. The finding that apperceptive processing difficulties in ASD can occur while associative processing is relatively intact is in line with previous findings on developmental phonagnosia (Roswandowitz et al., 2014; Roswandowitz et al., 2017). The ASD group showed impaired unfamiliar voice processing while the semantic association for familiar voices was intact (i.e. famous voice recognition). With regard to speech, our results indicate that speech processing is relatively intact (see study 2; Schelinski et al., 2014) however, reduced under noisy conditions (Schelinski & von Kriegstein, in prep.). This suggestion is based on additional results from a speech-in-noise recognition test that was performed by the same sample of participants. In this test, we investigated the individual thresholds for speech recognition when speech was presented with increasing noise levels. We found that individuals with ASD showed higher thresholds as compared to the control group, i.e. typically developed individuals understood speech in higher noise levels.

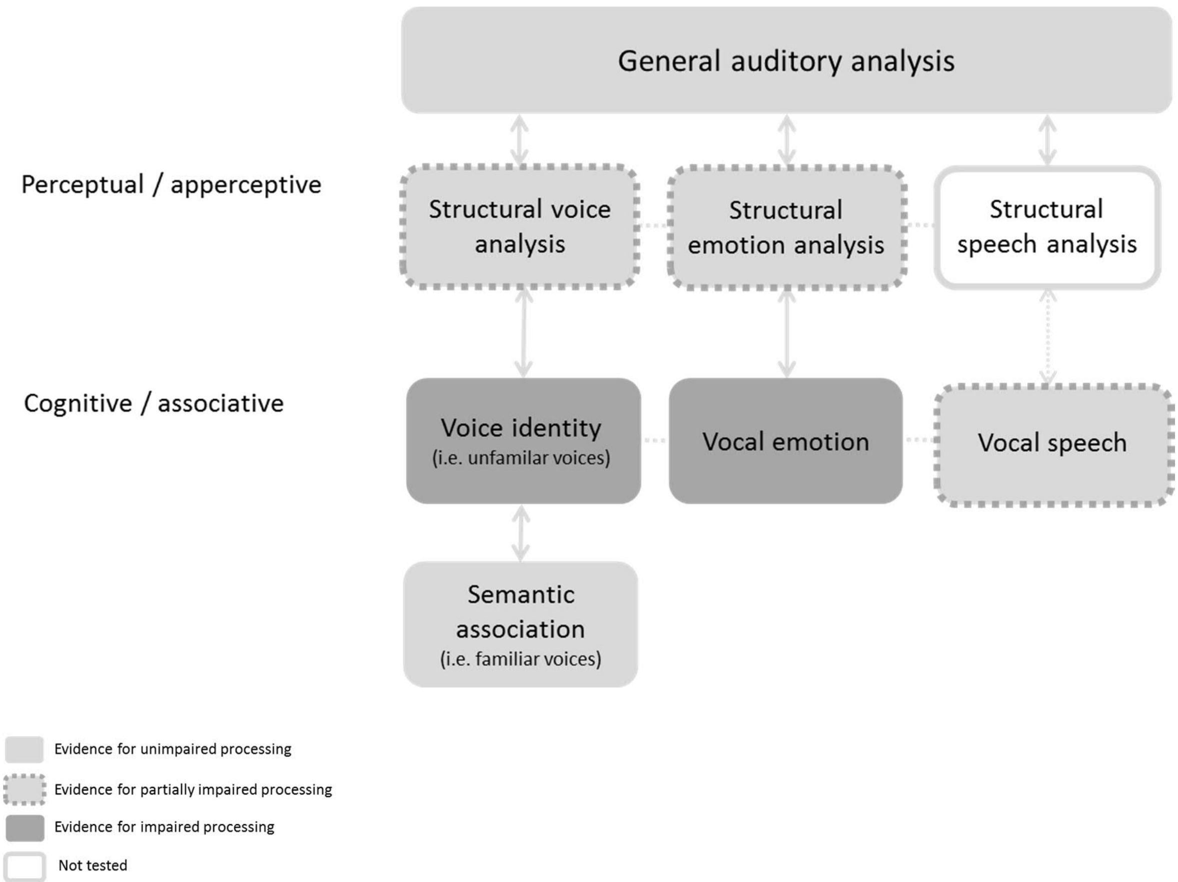

Figure 2
Working model on voice perception in ASD based on a recently introduced model on voice perception (Roswandowitz et al., under review; adapted from Belin et al., 2004; Blank et al., 2014; Ellis et al., 1997). The model is adapted based on results from the current work: Light grey boxes indicate evidence for unimpaired processing in high-functioning ASD. Light grey boxes with a dotted line indicate that we found evidence for partially impaired processing. Dark grey boxes indicate impaired processing in high-functioning ASD. Our study did not include a test on the structural speech analyses (white box), however our results suggest that speech recognition is relatively intact but impaired at higher noise levels. The results of the current dissertation suggest that voice perception impairments in ASD are of apperceptive nature, i.e. due to a dysfunction at an early stage of processing when the voice is encoded structurally. Behavioural and neuroimaging results merge to the view that impairments in the perception of acoustic voice features, such as vocal pitch perception might at least partly underlie impairments in higher cognitive voice processing, i.e. voice identity and vocal emotion perception.

The behavioural and neuroimaging results of the current work imply that perceptual impairments, i.e. impaired perception of vocal pitch might underlie deficits in socially relevant cognitive auditory processing, such as voice identity or vocal emotion recognition (Figure 2). In line with recent ASD theories the present work also supports the critical role of the STS/STG in explaining ASD symptomatology (see chapter 2.2). Perceptual impairments might contribute significantly to difficulties in social cognition (Gold et al., 2012). The current work suggests to consider impairments in perceptual processing as a potentially contributing variable to impairments in higher-cognitive processing in ASD and thus to the development of difficulties in social cognition and communication. This conclusion is in line with approaches that emphasise the importance of differences in lower-level sensory processing in explaining difficulties in social cognition in ASD (Baum et al., 2015; Dakin & Frith, 2005; Happe & Frith, 2006; Pellicano & Burr, 2012).

More generally, our results bring voice perception difficulties in ASD to the awareness of affected people, professionals and to the lay public. This might foster better understanding of the needs and difficulties people with ASD are faced with in everyday life. For example, difficulties in voice perception in ASD might be associated with avoidance of certain social situations such as phone calls, as it might be especially effortful for people with ASD in situations where only auditory information is available. Taking into account evidence that people with ASD also have difficulties in face recognition (for review see Weigelt, Koldewyn, & Kanwisher, 2012), person recognition difficulties in both the visual and auditory domain might exacerbate communication difficulties. Being aware of such impairments might lead to a better understanding of the difficulties encountered by people with ASD by the communication partner on one hand, and to a better self-awareness of one's own difficulties for individuals with ASD on the other hand. Being aware of difficulties in voice perception might be a first step towards the development of appropriate coping strategies to support communication.

4.3 Outlook

The current thesis opens up new avenues for future work on models on human communication as well as ASD research. In the following section I will highlight some of them.

The results of the current work open up new avenues for clinical application, i.e. diagnosing and intervention of ASD symptoms. For example, given that our results would be validated and thresholds for the difference between ASD and non-ASD participants would be quantified, testing impairments in perceptual processing of acoustic voice features, such as vocal pitch, might be a straightforward additional tool in the diagnostic procedure of ASD. Further, training of acoustic voice features might provide a potential remediation tool. In typically developed individuals, there is evidence that performance in basic auditory perception, such as non-vocal pitch perception, can be improved through training (Ari-Even Roth, Amir, Alaluf, Buchsenspanner, & Kishon-Rabin, 2003; Demany & Semal, 2002). Previous results suggest that training of non-vocal auditory perception, i.e. musical training, improves higher-level auditory vocal perception, i.e. speech-in-noise perception (Slater et al., 2015). Speculatively, such training effects might also be observed for the perception of vocal pitch and the training of vocal pitch might improve higher-level auditory vocal processing, such as vocal emotion or voice identity perception. However, the transfer of such training of vocal perceptual abilities to performance in higher-level vocal processing is currently unclear. The majority of previous studies on voice perception in ASD focused on single tasks in one modality (for review see Baum et al. 2015) and within single components of voice perception (see chapter 2.3). The results of the current work clearly suggest that there is a need to expand this approach by testing perceptual and cognitive abilities in ASD. Testing perceptual and higher cognitive abilities within the same sample of participants might for example contribute to clarifying the nature of the difficulties in higher-level processing that are part of the clinical picture of ASD.

ASD is a heterogeneous condition with different levels of severity and can be accompanied by intellectual and language impairment (APA, 2013). In our studies, individuals with high-functioning ASD participated. Replication studies representing the whole autism spectrum, i.e. big samples with individuals with several levels of symptom severity and examining different levels of cognitive and language function would increase the generalisability of the results and conclusions of the current work. This would also provide insights into the relation between difficulties in voice perception and the level of manifestation of ASD symptoms.

Investigating the developmental trajectory of voice perception impairments is another interesting direction for further studies. Previous studies showed that the ability to recognise the identity of another person by voice develops very early in development (e.g. De Casper & Fifer, 1980; Kisilevsky et al., 2003) and that voice-sensitive regions including the right STS/STG are functionally refined with increasing age (Bonte et al., 2016). ASD is a developmental disorder and symptoms typically manifest early in development (APA, 2013). There is evidence that atypical voice identity perception in ASD is also already present early in life (Klin, 1991). In the case of ASD it is unclear whether voice processing difficulties, such as impairments in voice identity or vocal emotion perception cause, exacerbate or originate from difficulties in social interaction. One possibility to test this would be to perform longitudinal studies on voice processing in ASD. Such a longitudinal approach would contribute to a better understanding of the causality between voice processing impairments and difficulties in social cognition and interaction in ASD. Longitudinal studies on voice perception in ASD would also provide more general insights into the developmental trajectory of voice perception impairments. So far, cases of developmental phonagnosia have been described for adults only (Garrido et al., 2009; Roswandowitz et al., 2014). However, the developmental trajectory of developmental voice identity processing impairments is currently unclear. ASD might be a good clinical condition to investigate the developmental trajectory of developmental voice identity processing impairments.

Another interesting approach relates to the comparison of voice perception in ASD and schizophrenia spectrum disorders (SSD; APA, 2013). Like ASD, SSD is conceptualised as a neurodevelopmental disorder that is associated with difficulties in social communication and interaction and there is an ongoing discussion about a partially shared etiology of these two conditions (for review see Chisholm, Lin, Abu-Akel, & Wood, 2015). Similar to ASD, SSD is associated with impaired voice identity and vocal emotion processing (for review see Conde, Goncalves, & Pinheiro, 2016). Interestingly, impairments in vocal emotion recognition in SSD have been suggested to be associated with altered processing of vocal pitch cues (Gold et al., 2012; Leitman et al., 2011). Additionally, as suggested for ASD in the current work, the altered processing of vocal emotion perception in SSD has been associated with dysfunction in the posterior STS/STG (Leitman et al., 2011). Future extended studies could systematically investigate and compare voice perception and its underlying mechanisms in individuals with ASD and SSD. A comparative approach on voice perception impairments and their underlying perceptual mechanisms in both conditions might contribute to a better understanding of their development and may provide further knowledge on the mechanisms and origins of voice perception impairments.

A range of further directions can be directly derived for research on voice perception from the current work. For example, it would be interesting to systematically investigate the neural mechanisms of impairments in vocal pitch processing and its relation to higher-level voice perception impairments, i.e. voice identity and vocal emotion perception. The current work also raises the question of shared and distinct processes in voice identity and vocal emotion perception. Previously, common and shared mechanisms between identity and emotion perception have been a topic in the face literature (Calder & Young, 2005). However, it would also contribute significantly to the understanding of the mechanisms of voice perception.

The current work also provides new approaches to systematically investigate face processing in ASD. Models on voice perception that guided the current work are based on models on face perception (see chapter 2.1.1) and there is good evidence for impaired face processing in ASD including impaired face identity (for reviews see Weigelt et al., 2012) and atypical facial emotion perception (for review see Harms, Martin, & Wallace, 2010). However, similar to the case of voice perception, the underlying mechanisms are currently unclear. For example, it is unclear, whether visual perception impairments also underlie often reported impaired face identity and facial emotion recognition. A similar systematic approach of underlying mechanisms of face perception as has been done for the perception of voices in the current work would provide evidence whether the dysfunctional processing mechanisms as suggested for voice perception also applies for the processing of faces. Such a similar approach would provide evidence whether the pattern is rather specific for auditory processing or directs to a more general processing style in ASD.

In more natural settings, communication does occur under more challenging listening conditions as compared to situations tested in the laboratory. In ASD there is evidence that vocal speech recognition is diminished under noisy conditions (Alcantara, Weisblatt, Moore, & Bolton, 2004). In another study which is not part of the current thesis, the same ASD group that showed intact vocal speech recognition when noise was not additionally superimposed to speech, also showed significantly worse performance, as compared to typically developed controls, in vocal speech recognition when noise was added to the speech stimuli (Schelinski & von Kriegstein, in prep.). In another ongoing study, which is also not part of the current thesis we investigated the neural mechanisms of this speech-in-noise processing deficit in ASD. Results from these studies will further contribute to a better understanding of the whole picture of voice processing in ASD under more natural, i.e. noisier conditions.

References

Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and Speech, 21*(4), 305-318.

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.

Alcantara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 45*(6), 1107-1114.

Andics, A., Gacsi, M., Farago, T., Kis, A., & Miklosi, A. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology, 24*(5), 574-578.

Andics, A., McQueen, J. M., Petersson, K. M., Gal, V., Rudas, G., & Vidnyanszky, Z. (2010). Neural mechanisms for voice recognition. *Neuroimage, 52*(4), 1528-1540.

Ari-Even Roth, D., Amir, O., Alaluf, L., Buchsenspanner, S., & Kishon-Rabin, L. (2003). The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. *Journal of Basic and Clinical Physiology and Pharmacology, 14*(2), 137-150.

Assal, G., Zander, E., Kremin, H. (1976), & Buttet, J. Discrimination des voix lors des lesions du cortex cerebral. *Archives Suisses de Neurologie, Neurochirurgie et de Psychiatrie, 119*(2), 307-315.

Assaneo, M. F., Sitt, J., Varoquaux, G., Sigman, M., Cohen, L., & Trevisan, M. A. (2016). Exploring the anatomical encoding of voice with a mathematical model of the vocal system. *Neuroimage, 141*, 31-39.

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (5th ed.). Washington, DC: American Psychiatric Association.

Barak, B., & Feng, G. (2016). Neurobiology of social behavior abnormalities in autism and Williams syndrome. *Nature Neuroscience, 19*(6), 647-655.

Baron-Cohen, S. (2008). *Autism and Asperger syndrome*. New York: Oxford University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163-175.

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.

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.

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine, 45*(3), 601-613.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129-135.

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience, 3*(10), 965-966.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Bestelmeyer, P. E. G., Belin, P., & Grosbras, M. H. (2011). Right temporal TMS impairs voice detection. *Current Biology, 21*(20), R838-R839.

Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *Journal of Neuroscience, 31*(36), 12906-12915.

Blank, H., Kiebel, S. J., & von Kriegstein, K. (2015). How the human brain exchanges information across sensory modalities to recognize other people. *Human Brain Mapping, 36*(1), 324-339.

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: merging evidence from patients and healthy individuals. *Neuroscience & Biobehavioral Reviews, 47C*, 717-734.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., . . . Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology, 21*(14), 1220-1224.

Bleuler, E. (1911). *Dementia Praecox oder Gruppe der Schizophrenien*. Leipizg: Deuticke.

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., . . . Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia, 48*(9), 2465-2475.

Bonte, M., Frost, M. A., Rutten, S., Ley, A., Formisano, E., & Goebel, R. (2013). Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage, 83*, 739-750.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience, 34*(13), 4548-4557.

Bonte, M., Ley, A., Scharke, W., & Formisano, E. (2016). Developmental refinement of cortical systems for speech and voice processing. *Neuroimage, 128*, 373-384.

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.

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of child psychology and psychiatry, and allied disciplines, 41*(7), 847-857.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Buchtel, H. A., & Stewart, J. D. (1989). Auditory agnosia - apperceptive or associative disorder. *Brain and Language, 37*(1), 12-25.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641-651.

Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S., . . . Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC neuroscience, 10*.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*(4), 231-239.

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.

Clopper, C. G., Rohrbeck, K. L., & Wagner, L. (2013). Perception of talker age by young adults with high-functioning autism. *Journal of Autism and Developmental Disorders, 43*(1), 134-146.

Conde, T., Goncalves, O. F., & Pinheiro, A. P. (2016). A cognitive neuroscience view of voice-processing abnormalities in schizophrenia: a window into auditory verbal hallucinations? *Harvard Review of Psychiatry, 24*(2), 148-163.

Crinion, J., & Price, C. J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain, 128*, 2858-2871.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron, 48*(3), 497-507.

Darwin, C. (2009). *The expression of the emotions in man and animals. (Original work published 1872)* (4th ed.). London: Harper-Collins.

De Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology, 3*(1), 69-86.

De Casper, A. J., & Fifer, W. P. (1980). Of human bonding - newborns prefer their mothers voices. *Science, 208*(4448), 1174-1176.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*(3), 1377-1388.

Domes, G., Heinrichs, M., Kumbier, E., Grossmann, A., Hauenstein, K., & Herpertz, S. C. (2013). Effects of intranasal oxytocin on the neural basis of face processing in autism spectrum disorder. *Biological Psychiatry, 74*(3), 164-171.

Dziobek, I., & Kohne, S. (2011). Brain imaging in autism spectrum disorders. *Nervenarzt, 82*(5), 564-572.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88*, 143-156.

Fairbanks, G., & Pronovost, W. (1938). Vocal pitch during simulated emotion. *Science, 88*, 382-383.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Paris: The Hague.

Fine, D. R. (2012). A life with prosopagnosia. *Cognitive Neuropsychology, 29*(5-6), 354-359.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*(3), 1511-1522.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research, 65*(6), 591-598.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "What"? Brain-based decoding of human voice and speech. *Science, 322*(5903), 970-973.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford: Blackwell.

Frith, U., & Happe, F. (1994). Autism: beyond "theory of mind". *Cognition, 50*(1-3), 115-132.

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews, 37*(1), 24-35.

Garcia-Villamisar, D., Rojahn, J., Zaja, R. H., & Jodra, M. (2010). Facial emotion processing and social adaptation in adults with and without autism spectrum disorder. *Research in Autism Spectrum Disorders, 4*(4), 755-762.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., . . . Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123-131.

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, Vols 1-5*, 152-155.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., I, S., . . . Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience, 7*(8), 801-802.

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.

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.

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. *American Journal of Psychiatry, 169*(4), 424-432.

Griffiths, T. D., & Hall, D. A. (2012). Mapping pitch representation in neural ensembles with fMRI. *Journal of Neuroscience, 32*(39), 13343-13347.

Groen, W. B., van Orsouw, L., Zwiers, M., Swinkels, S., van der Gaag, R. J., & Buitelaar, J. K. (2008). Gender in voice perception in autism. *Journal of Autism and Developmental Disorders, 38*(10), 1819-1826.

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.

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.

Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: a neuropsychological analysis. *Neuropsychologia, 48*(4), 1104-1114.

Hailstone, J. C., Ridgway, G. R., Bartlett, J. W., Goll, J. C., Buckley, A. H., Crutch, S. J., & Warren, J. D. (2011). Voice processing in dementia: a neuropsychological and neuroanatomical analysis. *Brain, 134*, 2535-2547.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America, 106*(2), 1064-1077.

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.

Happe, F., & Frith, U. (2014). Annual research review: towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry, 55*(6), 553-577.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290-322.

Harvey, P. D., & Penn, D. (2010). Social cognition: the key factor predicting social outcome in people with schizophrenia? *Psychiatry (Edgmont), 7*(2), 41-44.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Hertz-Picciotto, I., Croen, L. A., Hansen, R., Jones, C. R., van de Water, J., & Pessah, I. N. (2006). The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environ Health Perspect, 114*(7), 1119-1125.

Hickok, G., & Poeppel, D. (2007). Opinion - The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Hillier, A., Campbell, H., Keillor, J., Phillips, N., & Beversdorf, D. Q. (2007). Decreased false memory for visually presented shapes and symbols among adults on the autism spectrum. *Journal of clinical and experimental neuropsychology, 29*(6), 610-616.

Iarocci, G., Rombough, A., Yager, J., Weeks, D. J., & Chua, R. (2010). Visual influences on speech perception in children with autism. *Autism, 14*(4), 305-320.

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development, 82*(5), 1397-1403.

Ismail, M. M. T., Keynton, R. S., Mostapha, M. M. M. O., El Tanboly, A. H., Casanova, M. F., Gimel'farb, G. L., & El-Baz, A. (2016). Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey. *Frontiers in Human Neuroscience, 10*(211), 1-42.

Jarvinen-Pasley, A., & Heaton, P. (2007). Evidence for reduced domain-specificity in auditory processing in autism. *Developmental Science, 10*(6), 786-793.

Jiang, J., Liu, F., Wan, X., & Jiang, C. M. (2015). Perception of melodic contour and intonation in autism spectrum disorder: evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders, 45*(7), 2067-2075.

Jouventin, P., Aubin, T., & Lengagne, T. (1999). Finding a parent in a king penguin colony: the acoustic system of individual recognition. *Animal Behaviour, 57*, 1175-1183.

Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K., & Varma, S. (2012). Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity. *Neuroscience and Biobehavioral Reviews, 36*(4), 1292-1313.

Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. *Speech Separation by Humans and Machines*, 167-180.

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences, 16*(11), 559-572.

Kern, J. K., Trivedi, M. H., Grannemann, B. D., Garver, C. R., Johnson, D. G., Andrews, A. A., . . . Schroeder, J. L. (2007). Sensory correlations in autism. *Autism, 11*(2), 123-134.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., . . . Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry, 168*(9), 904-912.

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.

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.

Knapp, M., Romeo, R., & Beecham, J. (2009). Economic cost of autism in the UK. *Autism, 13*(3), 317-336.

Kreiman, J., Sidtis, D. (2011). *Foundations of voice studies*. Chichester, UK: Wiley-Blackwell.

Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *Neuroimage, 91C*, 375-385.

Lachs, L., & Pisoni, D. B. (2004). Specification of cross-modal source information in isolated kinematic displays of speech. *Journal of the Acoustical Society of America, 116*(1), 507-518.

Lang, C. J., Kneidl, O., Hielscher-Fastabend, M., & Heckmann, J. G. (2009). Voice recognition in aphasic and non-aphasic stroke patients. *Journal of Neurology, 256*(8), 1303-1306.

Lartseva, A., Dijkstra, T., & Buitelaar, J. K. (2015). Emotional language processing in autism spectrum disorders: a systematic review. *Frontiers in Human Neuroscience, 8*.

Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-induced changes in the cerebral processing of voice identity. *Cerebral Cortex, 21*(12), 2820-2828.

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.

Levey, D. J., Londono, G. A., Ungvari-Martin, J., Hiersoux, M. R., Jankowski, J. E., Poulsen, J. R., . . . Robinson, S. K. (2009). Urban mockingbirds quickly learn to identify individual humans. *Proceedings of the National Academy of Sciences of the United States of America, 106*(22), 8959-8962.

Lichtenstein, P., Carlstrom, E., Rastam, M., Gillberg, C., & Anckarsater, H. (2010). The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *The American journal of psychiatry, 167*(11), 1357-1363.

Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. *Archiv für Psychiatrie und Nervenkrankheiten, 21*, 222-270.

Magrath, R. D., Pitcher, B. J., & Gardner, J. L. (2009). Recognition of other species' aerial alarm calls: speaking the same language or learning another? *Proceedings of the Royal Society B-Biological Sciences, 276*(1657), 769-774.

Mathias, S. R., & von Kriegstein, K. (2014). How do we recognise who is speaking? *Frontiers in Bioscience, S6*, 92 - 109.

Maximo, J. O., Cadena, E. J., & Kana, R. K. (2014). The implications of brain connectivity in the neuropsychology of autism. *Neuropsychology Review, 24*(1), 16-31.

Mc Comb, K., Shannon, G., Sayialel, K. N., & Moss, C. (2014). Elephants can determine ethnicity, gender, and age from acoustic cues in human voices. *Proceedings of the National Academy of Sciences of the United States of America, 111*(14), 5433-5438.

Meister, H., Fursen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *Journal of Speech Language and Hearing Research, 59*(3), 546-556.

Mullennix, J. W., Johnson, K. A., Topcu Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98*(6), 3080-3095.

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*(6), 3219-3237.

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44*(3), 342-366.

Nickl-Jockschat, T., Habel, U., Michel, T. M., Manning, J., Laird, A. R., Fox, P. T., . . . Eickhoff, S. B. (2012). Brain structure anomalies in autism spectrum disorderua meta-analysis of VBM studies using anatomic likelihood estimation. *Human Brain Mapping, 33*(6), 1470-1489.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience and Biobehavioral Reviews, 36*(2), 836-854.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of child psychology and psychiatry, and allied disciplines, 32*(7), 1081-1105.

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.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504-510.

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research review: constraining heterogeneity: the social brain and its development in autism spectrum disorder. *Journal of child psychology and psychiatry, and allied disciplines, 52*(6), 631-644.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117 (Pt 6)*, 1283-1301.

Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia, 45*(8), 1899-1910.

Perrodin, C., Kayser, C., Abel, T. J., Logothetis, N. K., & Petkov, C. I. (2015). Who is that? Brain networks and mechanisms for identifying individuals. *Trends in Cognitive Sciences, 19*(12), 783-796.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience, 11*(3), 367-374.

Philip, R. C., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., . . . Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine, 40*(11), 1919-1929.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816-847.

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9*(2), 273-277.

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*(1), 236-250.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews, 32*(1), 123-142.

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2017). Emotional prosody processing in Autism Spectrum Disorder. *Social Cognitive and Affective Neuroscience*, *12*(2), 224-239.

Roswandowitz, C., Maguinness, C., & von Kriegstein, K. (under review). Deficits in voice-identity processing: acquired and developmental phonagnosia. In Süßholz S. & Belin P. (Eds.), *The Oxford Handbook of Voice Perception.*

Roswandowitz, 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.

Roswandowitz, C., Schelinski, S., & von Kriegstein, K. (2017). Distinct neural mechanisms in

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.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds - evidence from aphasia. *Brain, 126*, 928-945.

Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2013). Early auditory sensory processing of voices is facilitated by visual mechanisms. *Neuroimage, 77*, 237-245.

Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2015). Voice identity recognition: functional division of the right STS and its behavioral relevance. *Journal of Cognitive Neuroscience, 27*(2), 280-291.

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience, 11*(11), 1812-1822.

Schelinski, S., Riedel, P., & von Kriegstein, K. (2014). Visual abilities are important for auditory-only speech recognition: evidence from autism spectrum disorder. *Neuropsychologia, 65*, 1-11.

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research, 10*(1), 155-168.

Schelinski, S., & von Kriegstein, K. (in prep.). Speech-in-noise perception in adults with high-funcitoning autism spectrum disorder.

Schelinski, S., & von Kriegstein, K. (under review). Vocal emotion difficulties in autism spectrum disorder are associated with impaired vocal pitch processing.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123-148.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24-30.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zaske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews-Cognitive Science, 5*(1), 15-25.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., & Fink, G. R. (2001). The neural correlates of person familiarity - a functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804-815.

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behavioural Brain Research, 291*, 244-252.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*(5), 3177-3186.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America, 117*(1), 305-318.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of child psychology and psychiatry, and allied disciplines, 48*(8), 813-821.

Tick, B., Bolton, P., Happe, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, 57*(5), 585-595.

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-months-olds' social referencing behavior. *Developmental Science, 7*(3), 261-269.

Van Lancker, D. R., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829-834.

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.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia - a dissociation between familiar and unfamiliar voices. *Cortex, 24*(2), 195-209.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits - neuroanatomical correlates of phonagnosia. *Journal of clinical and experimental neuropsychology, 11*(5), 665-674.

von Kriegstein, K., Dogan, O., Gruter, M., Giraud, A. L., Kell, C. A., Gruter, T., . . . Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences of the United States of America, 105*(18), 6747-6752.

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.

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22*(2), 948-955.

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *Plos Biology, 4*(10), 1809-1820.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience, 17*(3), 367-376.

von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *Journal of Neuroscience, 30*(2), 629-638.

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior & Development, 6*(4), 491-498.

Waller, S. S., Eriksson, M., & Sorqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology, 6*(978), 1-11.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31*(3), 1389-1397.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience & Biobehavioral Reviews, 36*(3), 1060-1084.

Weintraub, K. (2011). Autism counts. *Nature, 479*(7371), 22-24.

World Health Organisation (2004). *International Statistical Classification of Diseases and Related Health Problems-10th Revision* (2 ed.). Geneva.

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(12), 2891-2902.

Xu, X., Biederman, I., Shilowich, B. E., Herald, S. B., Amir, O., & Allen, N. E. (2015). Developmental phonagnosia: neural correlates and a behavioral marker. *Brain and Language, 149*, 106-117.

Yang, D. Y. J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience and Biobehavioral Reviews, 51*, 263-275.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: a problem of recognition. *Journal of Psychosomatic Research, 65*(5), 445-451.

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974.

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences, 29*(7), 359-366.

Appendix

List of Figures

Selbstständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Stefanie Schelinski

Leipzig, den 18. Mai 2017