



Co-Editors: James W. Tanaka & Christian N.L. Olivers

Special issue:
Perceptual
Learning
Guest editors:
Gadi Yovel, Galia Avidan, and Alice O'Toole



ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: <http://www.tandfonline.com/loi/pvis20>

Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia

Corrina Maguinness & Katharina von Kriegstein

To cite this article: Corrina Maguinness & Katharina von Kriegstein (2017) Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia, *Visual Cognition*, 25:4-6, 644-657, DOI: [10.1080/13506285.2017.1313347](https://doi.org/10.1080/13506285.2017.1313347)

To link to this article: <https://doi.org/10.1080/13506285.2017.1313347>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 02 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 500



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia

Corrina Maguinness ^a and Katharina von Kriegstein ^{a,b}

^aMax Planck Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ^bDepartment of Psychology, Humboldt University of Berlin, Berlin, Germany

ABSTRACT

Conspicuous can be recognized from either the face or the voice alone. However, person identity information is rarely encountered in purely unimodal situations and there is increasing evidence that the face and voice interact in neurotypical identity processing. Conversely, developmental deficits have been observed that seem to be selective for face and voice recognition, developmental prosopagnosia and developmental phonagnosia, respectively. To date, studies on developmental prosopagnosia and phonagnosia have largely centred on within modality testing. Here, we review evidence from a small number of behavioural and neuroimaging studies which have examined the recognition of both faces *and* voices in these cohorts. A consensus from the findings is that, when tested in purely unimodal conditions, voice-identity processing appears normal in most cases of developmental prosopagnosia, as does face-identity processing in developmental phonagnosia. However, there is now first evidence that the multisensory nature of person identity impacts on identity recognition abilities in these cohorts. For example, unlike neurotypicals, auditory-only voice recognition is not enhanced in developmental prosopagnosia for voices which have been previously learned together with a face. This might also explain why the recognition of personally familiar voices is poorer in developmental prosopagnosics, compared to controls. In contrast, there is evidence that multisensory interactions might also lead to compensatory mechanisms in these disorders. For example, in developmental phonagnosia, voice recognition may be enhanced if voices have been learned with a corresponding face. Taken together, the reviewed findings challenge traditional models of person recognition which have assumed independence between face-identity and voice-identity processing and rather support an audio-visual model of human communication that assumes direction interactions between voice and face processing streams. In addition, the reviewed findings open up novel empirical research questions and have important implications for potential training regimes for developmental prosopagnosia and phonagnosia.

ARTICLE HISTORY

Received 24 August 2016
Accepted 22 March 2017

KEYWORDS

Phonagnosia; prosopagnosia;
cross-modal; multisensory;
voice-identity; face-identity

As humans, we can recognize the identity of a familiar person from either the face (Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowan, & Bruce, 1999) or the voice (Hanley, Smith, & Hadfield, 1998) alone. While for most of us these abilities emerge with ease, for some they can be severely limited. Developmental prosopagnosia is a person recognition disorder characterized by an inability to recognize identity from the face (Duchaine & Nakayama, 2005; Lee, Duchaine, Wilson, & Nakayama, 2010; McConachie, 1976; Susilo & Duchaine, 2013). Its auditory homologue, developmental phonagnosia, is characterized by a deficit in recognizing the identity of individual voices (Garrido et al., 2009; Herald, Xu, Biederman, Amir, & Shilowich, 2014; Roswandowitz et al., 2014). Both disorders appear to emerge in the

absence of brain insult (Duchaine & Nakayama, 2006b; Garrido et al., 2009; Roswandowitz et al., 2014).

It is estimated that approximately 2.5% of the population may have developmental prosopagnosia (Kennerknecht, Grueter, Welling, Wentzek, & Horst, 2006; Kennerknecht, Ho, & Wong, 2008). Recent prevalence estimates for the occurrence rate for developmental phonagnosia have been more varied. Some have suggested that it may reflect a rare disorder with anything within the range of 0.2% (Roswandowitz et al., 2014) to 1% (Xu, Biederman, Shilowich, Herald, & Amir, 2015) of the population affected. However, others have estimated a prevalence rate of 3.2% (Shilowich & Biederman, 2016), more within the lines of that reported for prosopagnosia. It is important to consider that self-awareness of a deficit in voice

CONTACT Corrina Maguinness  maguinness@cbs.mpg.de  Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

recognition may be less apparent than a deficit in face recognition, because other salient identity cues, such as the face, are often available during everyday interactions. Furthermore, developmental phonagnosia has been under examination for a much shorter time scale than developmental prosopagnosia, with the first case report on developmental prosopagnosia published in 1976 (McConachie, 1976) and on developmental phonagnosia in 2009 (Garrido et al., 2009). As such, it is possible that some initial estimates for developmental phonagnosia have been rather conservative.

Selective deficits in face-identity and voice-identity processing

Cases of developmental prosopagnosia and phonagnosia offer a unique opportunity to examine the perceptual and cognitive mechanisms involved in face and voice processing, and provide critical complementary results to examinations in neurotypical populations. To date, studies have largely focused on investigating the within modality specificity of each disorder. For example, examinations have centred on whether cases of developmental prosopagnosia are associated with a global deficit in face processing (e.g., facial affect and gender discrimination; Duchaine, Parker, & Nakayama, 2003; Le Grand et al., 2006; Nunn, Postma, & Pearson, 2001). Other studies have examined whether developmental prosopagnosia is associated with a broader impairment in visual within category discrimination (Dennett et al., 2012; Duchaine & Nakayama, 2005; Duchaine, Yovel, Butterworth, & Nakayama, 2006). These studies have highlighted that both developmental prosopagnosia and phonagnosia can,¹ respectively, present as selective deficits in the analysis of identity which are not mediated by a deficit in general face (for recent review see Susilo and Duchaine, 2013), voice (see Garrido et al., 2009; Roswadowitz et al., 2014), visual (Le Grand et al., 2006; for review see Kress and Daum, 2003) or auditory (Garrido et al., 2009; Roswadowitz et al., 2014) processing. However, few studies have examined cross-modal identity processing in prosopagnosia and phonagnosia.

Recent evidence suggests that voice-identity (Latinus & Belin, 2011; Zäske, Schweinberger, & Kawahara, 2010) may be represented in a similar manner to face-identity (Leopold, Rhodes, Müller, & Jeffery, 2005;

Valentine, 1991) in the neurotypical brain. For potentially facilitative purposes, these two identity sources appear to interact in person identity processing in neurotypical participants (Bülthoff & Newell, 2015; von Kriegstein et al., 2008; von Kriegstein & Giraud, 2006; O'Mahony & Newell, 2012; Schweinberger, Kloth, & Robertson, 2011). For example, listeners are more accurate at recognizing the identity of a speaker from the voice alone when the speaker has been previously learned by face, in comparison to auditory-only (Sheffert & Olson, 2004) or audio-visual control learning conditions. The audio-visual control conditions have included learning the speaker with their corresponding name (von Kriegstein & Giraud, 2006), or with a visual image depicting the occupation of the speaker (von Kriegstein et al., 2008; Schelinski, Riedel, & von Kriegstein, 2014; Schall, Kiebel, Maess, & von Kriegstein, 2013; for review see von Kriegstein, 2012). This behavioural improvement has been called the “face-benefit” (von Kriegstein et al., 2008) and it highlights that visual face processing mechanisms are behaviourally relevant for auditory-only voice-identity recognition.

These observations in neurotypicals generate questions about how audio-visual (face-voice) information may affect representations of vocal identity in cases of developmental prosopagnosia and in phonagnosia. In the pages that follow, we first review the nature of these audio-visual interactions in neurotypical processing, and then turn to examine the small number of behavioural and neuroimaging studies which have investigated voice-identity and face-identity processing in prosopagnosia and phonagnosia, respectively. We focus only on literature from group and case studies on phonagnosia and prosopagnosia which are of a *developmental* nature. Although we concentrate mainly on the impact of face information on voice processing, towards the end of the review we turn briefly to consider how voice information may affect the representation of facial identity in the neurotypical population and disorders of person recognition.

Voice-identity processing: audio-visual interactions in the neurotypical brain

When we listen to someone talking we often also see their corresponding face. Voices are therefore likely to become familiar within the context of their

corresponding facial identity and vice versa. Indeed, von Kriegstein, Kleinschmidt, Sterzer, and Giraud (2005) reported that auditory-only familiar voice recognition was associated with responses in voice-sensitive regions of the superior temporal sulcus/gyrus (STS/G) and responses in the fusiform face area (FFA). The FFA is a visual face-sensitive region (Kanwisher, McDermott, & Chun, 1997) in the fusiform gyrus which is involved in the processing of structural facial form and face-identity (Grill-Spector, Knouf, & Kanwisher, 2004; Liu, Harris, & Kanwisher, 2010; Kanwisher & Yovel, 2006; Axelrod & Yovel, 2015; Andrews & Ewbank, 2004; Weibert & Andrews, 2015; Ewbank & Andrews, 2008; Xu, Yue, Lescroart, Biederman, & Kim, 2009; Eger, Schyns, & Kleinschmidt, 2004).

Responses in the FFA appear to be behaviourally relevant for supporting auditory-only voice recognition (von Kriegstein et al., 2008; Schall et al., 2013). Specifically, the behavioural face-benefit is correlated with increased functional responses within the FFA (von Kriegstein et al., 2008). The FFA has also been shown to directly connect with voice-sensitive regions in the anterior/mid STS/G (Blank, Anwander, & von Kriegstein, 2011). These connections may provide a direct route for communication between these sensory regions and may facilitate responses in the FFA which have been reported as early as 110 ms after voice onset (Schall et al., 2013). Interactions between these sensory regions are consistent with the concordant properties of each stimulus which facilitate recognition (von Kriegstein et al., 2008). For example, static properties of the voice, such as vocal-tract resonance, map well to, and are predictive of, structural facial form (Ghazanfar et al., 2007; Krauss, Freyberg, & Morsella, 2002; Smith, Dunn, Baguley, & Stacey, 2016a; Smith, Dunn, Baguley, & Stacey, 2016b). These static vocal properties support voice-identity recognition (Lavner, Rosenhouse, & Gath, 2001). In contrast, more dynamic properties of the voice, such as formant transitions, map well to facial speech movement (for reviews see Campbell, 2008; Peelle and Sommers, 2015) and are used for speech recognition (Liberman, 1957; Sharf & Hemeyer, 1972; Smits, 2000; Benkí, 2003). In accordance, during voice recognition there are interactions between voice-sensitive areas and FFA, while during speech recognition there are interactions between speech areas and visual lip-movement areas in the left posterior superior temporal sulcus (pSTS) (von Kriegstein

et al., 2008; Schall & von Kriegstein, 2014). However, there is some emerging evidence that facial motion cues may also assist in preserving auditory voice recognition in noise, as listeners attend to more dynamic aspects of the voice, such as articulatory pattern, to support recognition (Maguinness & von Kriegstein, 2016). Under these degraded listening conditions, functional connections between the facial motion-sensitive right pSTS and voice-sensitive regions have been observed (Maguinness & von Kriegstein, 2016). See Yovel and O'Toole (2016) for a recent review on how dynamic aspects of the face and voice may support person recognition.

These observations provide strong support for an audio-visual model of human communication (von Kriegstein et al., 2008), which proposes that the brain uses previously encoded facial cues to assist in predicting, and thus enhancing, the recognition of the incoming vocal signal (see Figure 1 for an overview of face-voice interactions in neurotypical and atypical identity processing). Importantly responses in the FFA appear to support the perceptual processing of voices (Blank, Kiebel, & von Kriegstein, 2015; Schall et al., 2013), providing evidence that the face and voice interact to support identity processing at earlier stages of processing than previously assumed (Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Ellis, Jones, & Mosdell, 1997). More traditional models of person recognition propose that the face and voice undergo extensive unisensory processing and only interact to support recognition at supramodal, i.e., post-perceptual, stages of processing (Burton et al., 1990; Bruce & Young, 1986; see Blank, Wieland, and von Kriegstein, 2014; Barton and Corrow, 2016; Quaranta et al., 2016, for more recent reviews of these models).

Developmental prosopagnosia

Voice-identity processing in developmental prosopagnosia: behavioural investigations

A number of studies which examined cases of prosopagnosia documented that such individuals rely on alternatives cues, including vocal information, to identify others (Grueter, Grueter, Bell, Horst, & Laskowski, 2007; Kennerknecht et al., 2006; Kennerknecht, Kischka, Stemper, Elze, & Stollhoff, 2011; Le Grand et al., 2006; McConachie, 1976). However, such reports are largely based on anecdotal evidence.

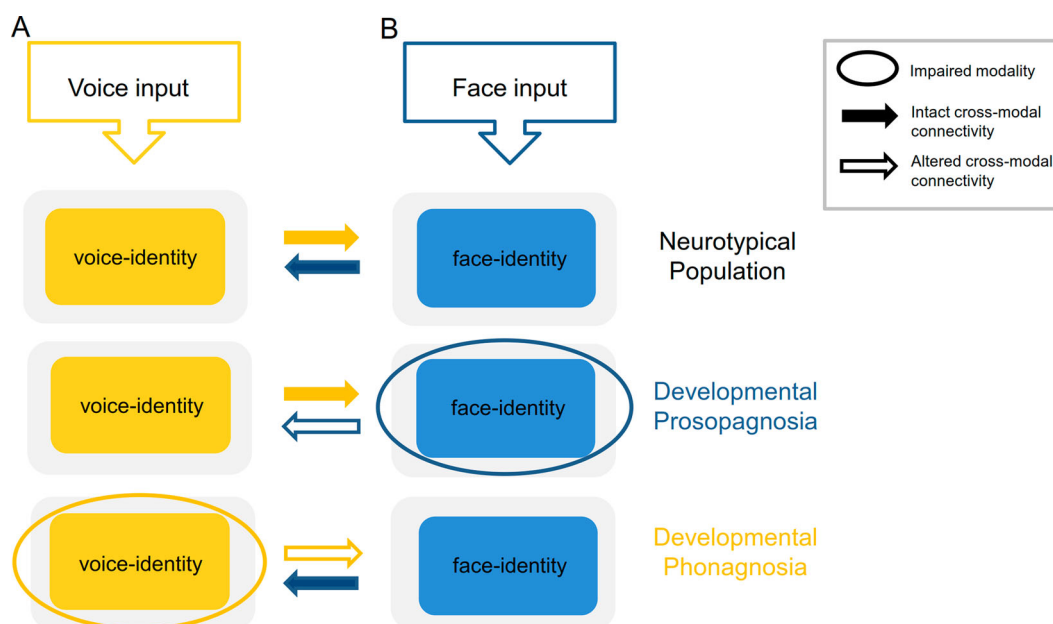


Figure 1. A schematic representation of the interactions between voice and face information, during identity processing based on unisensory input, in the neurotypical population, developmental prosopagnosia, and developmental phonagnosia. **Neurotypical population.** Auditory-only voice-identity processing (A) is facilitated by visual face-identity information in the FFA. The FFA shares connections with voice-sensitive regions in the a/m STS/G (solid yellow and blue arrows) (Blank et al., 2011; von Kriegstein et al., 2008; Schall et al., 2013). Although speculative, visual face processing (B) may also be facilitated by connections between these regions (see Bühlhoff and Newell, 2015, for behavioural effects). Although the solid yellow and blue arrows are depicted here with equal strength, it is likely that the actual strength of the interactions will vary depending on the saliency of the unisensory input. **Developmental prosopagnosia.** Auditory-only voice-identity processing (A) does not benefit from prior face-voice learning owing to impaired face-identity processing. Although connections between the FFA and STS/G exist in this cohort (von Kriegstein et al., 2006; Schall & von Kriegstein, 2014), they are not sufficient to optimise speaker recognition. This is likely owing to atypical recruitment of the FFA during voice-identity processing which may alter the nature of the information transferred between these regions (outline blue arrow) (von Kriegstein et al., 2006, 2008). Hypothetically, visual face processing (B) could be enhanced through audio-visual face-voice learning in this cohort (solid yellow arrow). **Developmental phonagnosia.** Auditory-only voice-identity processing (A) may be enhanced through compensatory recruitment of intact visual face-identity mechanisms (solid blue arrow) (Roswadowsitz et al., 2014; Roswadowsitz et al., 2017). However, visual face processing (B) may not benefit from additional vocal information in this cohort, owing to the failure to represent the voice at the level of the individual identity (outline yellow arrow).

One might predict that voice recognition in prosopagnosics may be superior to neurotypicals. This might be in a similar vein to blind individuals, who exhibit remarkable compensatory abilities for supporting person recognition via the auditory modality (Bull, Rathborn, & Clifford, 1983; Föcker, Best, Hölig, & Röder, 2012; Föcker, Hölig, Best, & Röder, 2015; Holig, Focker, Best, Roder, & Buchel, 2014b). However, this hypothesis has been shown to be incorrect, as evidenced from the findings of the handful of behavioural studies which have explicitly examined voice recognition in developmental prosopagnosia (Jones & Tranel, 2001; von Kriegstein, Kleinschmidt, & Giraud, 2006; von Kriegstein et al., 2008; Liu, Corrow, Pancaroglu, Duchaine, & Barton, 2015). These studies are reviewed in detail below.

Jones and Tranel (2001) examined voice processing in a single case of prosopagnosia. They reported the

case of TA, a 5-year-old boy who presented with a marked impairment in visual face recognition. So severe was his face processing deficit that his parents and teachers had observed the child approach complete strangers and ask if they were his father. However, at least anecdotally, TA appeared to show a preserved ability to recognize people, including his parents, by their voice. Jones and Tranel (2001) explicitly examined TA's voice recognition in a task which required naming the identity of a series of familiar voices, articulating a standard sentence. TA's performance was poorer than his two age-matched controls. He accurately recognized 5/8 identities, while his controls performed at ceiling recognizing 6/6 identities. The full details of the vocal identities presented to TA and controls are not detailed by the authors, although extrapolation from their familiar face recognition test design would suggest that the voices were

that of friends, family, and schoolmates. Although impaired, TA's familiar voice recognition performance was still superior to his familiar face recognition performance (Jones & Tranel, 2001).

von Kriegstein et al. (2006) also assessed voice processing in an additional single case of prosopagnosia. They examined familiar and unfamiliar voice recognition in the case of SO, a 35-year-old female who presented with a deficit in face recognition, in spite of normal general visual processing skills. SO reported that she relied on the voice as a cue to recognize others. In that study, both SO and a group of control participants ($n = 9$) were presented with a target familiar or unfamiliar voice, and were asked to indicate when they heard this target repeated within a series of other vocal identities. In a control condition, participants were presented with the same stimulus material (i.e., familiar or unfamiliar speakers), but were presented with target words and asked to indicate if the words were repeated within the series of voice samples. Familiar speakers were personal acquaintances who were encountered regularly on a day-to-day basis. von Kriegstein et al. (2006) observed that SO's ability to recognize familiar speakers was significantly impaired. SO performed at 69% correct for familiar voice recognition trials, while her controls performed at 99% correct. Interestingly, her unfamiliar voice recognition performance was not significantly different from controls (SO = 71% correct; controls = 86% correct). SO also showed a preserved ability to recognize the verbal content of sentences, spoken by both familiar and unfamiliar speakers. As such, SO's voice-identity recognition deficit could not be explained in terms of a *general* deficit in voice recognition or auditory processing. Rather, SO's impairment appeared to be restricted to the recognition of the familiar voices of the people that she interacted with daily.

von Kriegstein et al. (2008) went on to examine the hypothesis that the impaired processing of face-identity in prosopagnosia may affect voice-identity processing only for voices that have been encountered together with a face, a situation which is common for our encounters with personally familiar people's voices. They used a paradigm which directly manipulated the availability of face information during voice-identity learning. In that study, a sample of prosopagnosics ($n = 17$) and control participants ($n = 17$) learned a series of unfamiliar voices with their

corresponding face, or a visual control image which depicted the occupation of the speaker. The authors noted that the control group showed a significant face-benefit on subsequent auditory-only voice recognition (mean face-benefit 5.27%). In contrast, the prosopagnosic group failed to demonstrate this behavioural enhancement, demonstrating a face-benefit of -1.81%. Importantly, prosopagnosics auditory-only recognition for voices learned in the visual control condition was similar to controls.

Unimpaired auditory-only voice recognition for voices learned without a face has also been replicated in a further study of prosopagnosics ($n = 12$) and controls ($n = 73$) (Liu et al., 2015). In that study, listeners were asked to learn a series of unfamiliar voice identities presented in blocks. In each block, participants learned the identity of three voices. Immediately following learning, participants were presented with a series of voice pairs; each containing a learned and a novel identity. Listeners indicated which of the voices matched the voice from the learning phase. Liu et al. noted that all of the prosopagnosic cases had comparable performance levels to controls on this task. In the voice discrimination task, which involved a delayed match to sample two alternative forced choice design on unfamiliar voices, all but one case of prosopagnosia, the case of DP035 a 40-year-old male, showed normal performance. DP035 also scored poorly on the Cambridge Face Perception Test, a standardized measure of face perception ability (Duchaine, Germine, & Nakayama, 2007). The authors suggested that DP035 may have separate apperceptive (i.e., impaired perceptual feature analysis) deficits in face (De Renzi, Faglioni, Grossi, & Nichelli, 1991) and voice (Roswadowitz et al., 2014) processing, rather than a multimodal recognition deficit.

Taken together, the results show that voice recognition in prosopagnosics is similar, but not superior,² to neurotypical controls if voices are learned in the auditory modality only (Liu et al., 2015) or in multisensory conditions, but without a face (von Kriegstein et al., 2008). In contrast, familiar voices appear to be more poorly represented in prosopagnosics compared to controls (Jones & Tranel, 2001; von Kriegstein et al., 2006). Familiar voices are often learned within the presence of their corresponding facial identity. This is particularly true for personally familiar voices (i.e., the voices used in von Kriegstein et al.,

2006, and most likely in Jones and Tranel, 2001), where regular face-to-face interactions are typical. The findings of impaired voice recognition in this context, suggests that it may be the atypical processing of face information in prosopagnosia which impacts on the ability to robustly represent the identity of familiar voices, i.e., altered cross-modal interactions between face-identity and voice-identity cues. In the following section, we report on the neuroimaging studies which have complemented these behavioural investigations.

Voice-identity processing in developmental prosopagnosia: neuroimaging investigations on audio-visual interactions

In the case of SO, von Kriegstein et al. (2006) also used functional magnetic resonance imaging (fMRI) to examine responses while SO and controls recognized familiar and unfamiliar voices (task as described in behavioural investigations section above). They noted that the response in the FFA to familiar voices was preserved in SO, in spite of her poor familiar voice recognition performance. Furthermore, normal functional connectivity between the FFA and the voice-sensitive STS/G was observed in SO. In contrast, she had reduced functional connectivity between the FFA and the extended system. The extended system included supramodal regions such as the anterior temporal lobe, which respond to familiar voice and face stimuli, independent of the sensory input modality (Avidan & Behrmann, 2014; Blank et al., 2014; Haxby, Hoffman, & Gobbini, 2000). However, the authors argued that the observed response in the FFA may not have been sufficient to enhance familiar voice recognition. Specifically, they proposed that abnormal responses in the FFA in prosopagnosia to faces at the individual (e.g., Schiltz et al., 2006), rather than at a categorical (e.g., Rivolta et al., 2014), level may have impacted on SO's ability to benefit from multi-modal representations of identities.

The preserved responses in the FFA, and the preserved connectivity between the FFA and the voice-sensitive STS/G, support the argument that face and voice interactions in identity processing occur at early perceptual (Schall et al., 2013; for review see von Kriegstein, 2012), rather than late post-perceptual (Bruce & Young, 1986; Burton et al., 1990; Ellis et al., 1997), stages of identity processing. Note that SO's

reduced connectivity between the FFA and extended system implies that FFA responses during voice recognition are unlikely to be driven by top-down modulation from supramodal brain regions, which would be assumed by more traditional models of person recognition (Bruce & Young, 1986; Burton et al., 1990; Ellis et al., 1997).

von Kriegstein et al. (2008) also examined functional responses in the FFA in a larger group of prosopagnosics ($n = 17$) and controls ($n = 17$), during the recognition of speakers who had been previously learned by face, or with a visual control image (see preceding behavioural investigations section for full details of the task design). They observed increased functional responses in the FFA for both prosopagnosics and controls during the auditory-only recognition of face-learned speakers. Such a finding parallels the preserved responses in the FFA to familiar voices in the case of SO (von Kriegstein et al., 2006). Crucially a correlation between the magnitude of these FFA responses and the behavioural face-benefit score was evident in the control group only. In line with von Kriegstein et al., 2006, this suggests that prosopagnosics do not recruit the FFA to *optimize* voice recognition and that it is therefore the integrity of the neural processing within the FFA which is altered in cases of prosopagnosia, rather than overall magnitude of responses within the region itself. Schall and von Kriegstein (2014) later went on to reveal normal functional connectivity between the FFA and the voice-sensitive STS/G in prosopagnosics ($n = 17$) on the same dataset, corroborating the findings from SO, and again highlighting that interactions between these sensory cortices likely occur at a perceptual stage of processing. These findings support an audio-visual model of human communication (von Kriegstein et al., 2008) (see Figure 1), as they demonstrate that atypical processing in one modality, e.g., failure to represent facial identity at an individual level may modulate voice-identity processing abilities.

On a side note, it is also intriguing to note that a number of studies have also observed responses in fusiform areas, in response to vocal input in both congenitally (i.e., blind since birth) (Gougoux et al., 2009; Holig, Focker, Best, Roder, & Buchel, 2014a) and late blind (i.e., blindness acquired post birth) (Holig et al., 2014b) individuals. It has been speculated that these fusiform responses may arise as blindness leads to the strengthening of *pre-existing* connections

between face-sensitive and voice-sensitive regions (Holig et al., 2014a, 2014b). Such strengthening of connections between cortical regions may lead to a partial transfer of voice-identity processing from the STS/G to the fusiform gyrus (Holig et al., 2014b, 2014a). Such observations highlight how fundamental these audio-visual connections may be in the brain.

Developmental phonagnosia

Face-identity processing in developmental phonagnosia: behavioural investigations

To date, a total of five cases of developmental phonagnosia have been reported in the literature (Herald et al., 2014; Garrido et al., 2009; Roswadowitz et al., 2014). The majority of these published cases, four out of five (exception is the case of SR, briefly reported in Xu et al., 2015), have included assessment of face-identity processing abilities. One view could be that phonagnosics would be better than neurotypicals at face-recognition as a potential compensatory mechanism for their phonagnosia. For example, there is some evidence that deaf individuals may show enhanced face-identity processing (e.g., Arnold & Murray, 1998; Bellugi et al., 1990; but see also Arnold and Mills, 2001; McCullough and Emmorey, 1997, for mixed findings). The alternative view is that, similar to prosopagnosics, there are no compensatory mechanisms in the other modality. So far this alternative view is supported by the studies, which we review below.

In the first documented case of developmental phonagnosia, Garrido et al. (2009) examined the face recognition abilities of KH, a 60-year-old female with a lifelong impairment in voice recognition, using the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006a) and the Famous Faces Test (FFT; Duchaine & Nakayama, 2005). The CFMT is a standardized test which examines face recognition performance using recently learned unfamiliar faces. The FFT examines familiar face processing abilities. KH's performance on the CFMT was 67 out of 72, while her controls ($n = 7$) scored on average 57.9 out of 72 ($SD = 7.9$). Although, KH's score was above the controls' average, there was no evidence that KH and the control group's performances were statistically different on this task, $p = .35$; statistical differences assessed by current authors using a modified

t -test (two-tailed probability) for single case analysis (Crawford & Howell, 1998). The same pattern is evident for a one-tailed probability analysis. Similarly, her performance on the FFT was within the normal range of control participants ($n = 5$). A similar performance profile of intact face recognition was observed for the case of AN, a 20-year-old female phonagnosic (Herald et al., 2014; Xu et al., 2015). AN had normal performance on two similarly designed tests, based on the FFT, which examined familiar face recognition (FFT; Duchaine & Nakayama, 2005). Intriguingly, AN stated that she was not aware of her voice-identity deficit growing up as she had not considered that people could recognize a person *without* seeing their face.

Roswadowitz et al. (2014) also noted intact face recognition abilities in two cases of developmental phonagnosia; AS and SP. AS was a 32-year-old female, who was characterized as having an apperceptive phonagnosia (i.e., poor perceptual analysis of voice-identity cues) following a detailed battery of voice and auditory processing tests. Anecdotally, AS reported that she found it hard to discriminate the voice of her daughter, from her daughter's friend, when they were playing in a nearby room. SP was a 32-year-old male who, like AS, reported a deficit in recognizing identity from the voice alone. In contrast to AS, SP was characterized as having an associative variant of phonagnosia (i.e., a deficit in semantic association to the voice-identity, as opposed to poor perceptual analysis of voice individuating properties). Both cases performed similarly to their respective controls on the CFMT. In addition, both AS and SP had normal performance on a task which required them to associate and recognize unfamiliar face-name combinations (face-name test; Roswadowitz et al., 2014).

Voice-identity processing in developmental phonagnosia: behavioural investigations on audio-visual interactions

The studies above suggest normal face-processing in cases of phonagnosia. However, studies examining how the face and voice may interact in identity processing in this population are scarce. One study to examine such interactions was Roswadowitz et al. (2014). In that study, subjects were required to learn to associate unfamiliar voices with facial identities. Following learning, participants listened to each learned

voice and chose which facial identity, from three presented images, was associated with the corresponding voice-identity. Intriguingly, relative to her controls, AS (apperceptive phonagnosic) showed only a trend for impaired performance on the unfamiliar voice-face learning task (AS = 73% correct; controls ($n = 11$) = 87%). In marked contrast, her ability to link voice identities with colours (AS = 47% correct; controls ($n = 11$) = 74%) was significantly impaired. AS's impairment in linking voices to colours is not surprising given her poor ability to learn vocal identities. However, her somewhat preserved ability to link voices to faces is suggestive that AS may have had a relatively preserved ability to use additional facial information to enhance the representation of the vocal percept. Nevertheless, it is important to note there was no evidence for a classical dissociation in her performance between the two tasks ($p = .15$; assessed by current authors using the Bayesian Standardised Difference Test, see Crawford and Garthwaite (2007)). This highlights that although AS's representation of vocal identities may benefit, to some extent, from additional facial information, her performance is still not on par with her controls. AS was also significantly impaired on a separate task which required linking voices with names (AS = 50% correct; controls ($n = 343$) = 75%). On the other hand, SP (associative phonagnosic) was impaired, compared to controls, on all tasks which required linking voices to visual stimuli (faces, colours, names).

Voice-identity processing in developmental phonagnosia: neuroimaging investigations on audio-visual interactions

In a follow up study, Roswadowitz and colleagues investigated the neurological underpinning of AS and SP's selective voice-identity processing deficits (Roswadowitz, Schelinski, & von Kriegstein, 2017). In that study, they used fMRI to examine the responses in auditory and visual regions (among others) as both phonagnosics and controls recognized the identity of a series of speakers (speaker task). In a control condition, both phonagnosics and controls recognized the speech content (speech task) on the same auditory stimuli. The contrast of speaker, compared to speech task, has been shown to reveal brain regions involved in voice-identity processing (von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003).

Relative to controls ($n = 16$), AS had reduced responses in regions of the core-voice system which support the perceptual analysis of vocal identity. These regions included the right antero-lateral Heschl's gyrus and planum temporale and the right posterior STS/G. On the other hand, relative to controls ($n = 16$), SP showed enhanced responses in the right posterior STS/G of the core-voice system, but reduced connectivity between the core-voice and extended system. These neural profiles fit well with AS and SP's respective apperceptive and associative phonagnosias (see Roswadowitz, Maguinness, and von Kriegstein, *in press*, for a recent overview of the core-voice and extended system and how it relates to subtypes of developmental phonagnosia).

Interestingly, Roswadowitz et al. (2017) also found a trend for increased functional responses in the FFA in AS compared to controls, during speaker (i.e., voice recognition) compared to speech recognition. SP did not show these same cross-modal responses to vocal input. It is possible that AS may use supplementary facial information to enhance her weak perceptual processing of voices. The trend for increased responses in the FFA during auditory-only voice recognition may be reflective of this cross-modal compensation (von Kriegstein et al., 2006, 2008). In contrast, as SP's perceptual processing of voice stimuli was intact (note SP is an associative phonagnosic), he may not rely so heavily on this compensatory perceptual mechanism. However, it is important to note that the auditory voices presented to AS and SP during functional image acquisition had *not* been previously learned by face. Moreover, AS's voice recognition performance still remained poor (AS = 44.91%, controls = 86.86%) in spite of these cross-modal responses. However, it is plausible that these responses are reflective of the weighting that AS gives to faces during the encoding of voice-identity (note AS has a relatively preserved ability to link faces and voices).

Face-identity processing: audio-visual interactions

Can audio-visual learning enhance face recognition in the general population?

In general, it has been proposed in multisensory research that the benefit of multisensory presentation is greatest when the reliability of one (or more) of the

sensory cues is low (Meredith & Stein, 1986; for review see Alais, Newell, & Mamassian, 2010). In addition, the most reliable signal is weighted more heavily by the perceptual system (i.e., maximum likelihood model; see Alais & Burr, 2004; Alais et al., 2010; Ernst & Banks, 2002). The face has been shown to be a more reliable cue to identity than the voice (Joassin et al., 2011; Stevenage, Hale, Morgan, & Neil, 2014; Stevenage, Hugill, & Lewis, 2012). Such a mechanism may therefore explain the face-benefit for voice-identity recognition (von Kriegstein et al., 2008), where facial cues support the representation of voice-identity.³ Interestingly, there is recent behavioural evidence that audio-visual interactions may also support face, in addition to voice, recognition (Bülthoff & Newell, 2015; see also Bülthoff & Newell, 2017). Specifically, Bülthoff and Newell (2015) noted that visual-only face recognition was more accurate for faces that had been previously learned with distinctive, in comparison to typical, voices. The same effect was not observed when faces were paired with distinctive sounds (musical chords) during audio-visual learning. Such a finding suggests it was the naturalistic coupling of the face with vocal, rather than arbitrary auditory information, which mediated the behavioural enhancement on face recognition. This provides the first evidence that the perceptual system may use previously learned vocal cues to modulate the subsequent visual-only recognition of faces. The effect could be conceived as a “voice-benefit” on face recognition.

Because the face is a more salient cue to identity than the voice, it is unlikely that the cross-modal modulatory effects observed in voice recognition (von Kriegstein et al., 2008) would be entirely equitable for the recognition of faces. It is plausible that such interactions may assist in recognizing a face under situations of visual uncertainty, e.g., in suboptimal viewing conditions (Joassin, Maurage, & Campanella, 2008), or when the face is typical in appearance (Valentine, 1991). While the neurological underpinnings of the effects described by Bülthoff and Newell (2015) are currently unknown, we cautiously propose that these effects could be interpreted within the framework of an audio-visual model of human communication (von Kriegstein, 2012; von Kriegstein et al., 2008). For example, it is conceivable that under certain circumstances responses in voice-sensitive regions of the STS/G may be observed

during the recognition of faces learned by voice. This may be governed by direct connections between face-sensitive and voice-sensitive regions (Blank et al., 2011), which facilitate the sharing of identity information across sensory cortices.

Face-identity processing in prosopagnosia and phonagnosia: open questions on audio-visual interactions

There are first indications that intact face-identity processing might enhance voice-identity representations in phonagnosia (case of AS; Roswandowitz et al., 2014), a disorder which is characterized by the impaired processing of vocal identity. This is suggestive of cross-modal compensation: intact processing in one modality, i.e., visual face processing, may be able to help bootstrap processing in the impaired modality, i.e., auditory voice processing. However, it is important to note that while this compensation is beneficial, it is not sufficient to produce *neurotypical* levels of voice processing performance (case of AS). Phonagnosics also, by definition, show poor recognition of familiar voices which are usually encountered in a natural audio-visual learning environment. Examining whether responses in the FFA may underpin a potential face-benefit on voice-processing in phonagnosia, and examining the integrity of these responses during familiar voice processing, is therefore warranted.

The observation that face-identity cues may assist in enhancing voice processing in phonagnosia (Roswandowitz et al., 2014) and that voice-identity cues may, under certain circumstances, modulate face-identity representations in neurotypicals (Bülthoff & Newell, 2015), raises some potentially fruitful questions for prosopagnosia research. Specifically, can prosopagnosics use vocal information to enhance weak face-identity representations? To date, attempts to improve face-identity processing in this cohort have focused on the use of visual-only training paradigms (for recent reviews see Bate and Bennetts, 2014; DeGutis, Chiu, Grosso, and Cohan, 2014). It is possible that training paradigms which include additional vocal information may be of benefit in this cohort (see Figure 1).

Future considerations may also explore how impaired voice-identity processing, as evidenced in phonagnosia, may impact on face-identity processing within the context of audio-visual learning (see

Bülthoff & Newell, 2015). One can speculate that, in parallel to prosopagnosics (von Kriegstein et al., 2006, 2008; Schall & von Kriegstein, 2014), failure to represent vocal identities at the individual level may mitigate any potential voice-benefit on face processing (see Figure 1). Though, it would also be important to further elucidate under what specific circumstances face processing may benefit from additional voice-identity cues (e.g., possibly under situations of visual uncertainty) in neurotypical processing.

With the above in mind, we propose some potential questions for consideration in future research. Under what circumstances can audio-visual learning support the unimodal recognition of voices and faces in neurotypical processing? Can prosopagnosics use additional voice-identity cues to improve face-identity processing? What type of information is transferred between the FFA and the a/m STS/G during voice-identity processing in developmental prosopagnosics? Can phonagnosics compensate for their voice recognition deficit via recruitment of visual face mechanisms, e.g., in the FFA?

Conclusion

The reviewed findings demonstrate that cross-modal processing of the face and voice can be altered in cases of prosopagnosia and phonagnosia. In comparison to neurotypicals, prosopagnosics show poorer recognition of both familiar voices and voices that have been recently learned by face (von Kriegstein et al., 2006, 2008). This poor performance profile is mediated by deficits in using facial identity cues to enhance voice representations (von Kriegstein et al., 2006, 2008) and is associated with atypical recruitment of the FFA during voice recognition (von Kriegstein et al., 2008). This highlights that impaired processing in one modality, i.e., visual face processing, can directly impact on the ability to represent identity in another modality i.e., auditory voice processing. Importantly, the unimodal representation of voice identities in prosopagnosics appears to be largely comparable to controls (von Kriegstein et al., 2008; Liu et al., 2015). However, in everyday life voices are rarely learned in a unimodal fashion (except maybe when interacting with call centres or when listening to the radio); therefore, it is likely that impaired face-processing

may have a real impact on prosopagnosics day-to-day voice-recognition experiences.

Furthermore, a consistent finding from the reviewed studies in both neurotypicals and in disorders of person recognition is that interactions between the face and the voice emerge at early, i.e., perceptual rather than late stages of identity processing (von Kriegstein et al., 2006; Schall et al., 2013). Future studies may concentrate on how these cross-modal interactions are shaped in different subtypes of prosopagnosia and phonagnosia, i.e., apperceptive versus associative. It is conceivable, and there is some evidence to suggest (Roswandowitz et al., 2014; Roswandowitz et al., 2017), that the addition of cross-modal identity cues may be particularly facilitative in boosting the perceptual processing of identity in those who present with an apperceptive, rather than associative, variant of prosopagnosia or phonagnosia.

Taken together, the reviewed findings challenge traditional models of person recognition which have proposed an independence between face-identity and voice-identity processing (Bruce & Young, 1986; Burton et al., 1990; Ellis et al., 1997) and instead support an audio-visual model of human communication with direct interactions between face-identity and voice-identity processing streams (von Kriegstein et al., 2005; von Kriegstein et al., 2008; reviewed in von Kriegstein, 2012). This audio-visual model in turn proposes some potentially fruitful avenues for future research, for example, exploring the impact of audio-visual perceptual training on identity processing in prosopagnosia and phonagnosia.

Notes

1. The findings are not homogenous across all cases of prosopagnosia, for overviews see Susilo and Duchaine (2013); Behrmann and Avidan (2005).
2. A reviewer alerted us to the possibility that some undiagnosed cases of prosopagnosia may show superior voice recognition abilities. Conceivably, such individuals may be less likely to come forward for diagnosis of prosopagnosia, as they already compensate well for their disorder via auditory recognition cues. However, such a cohort has not been revealed yet.
3. It has also been reported that voice recognition can be impaired, rather than improved, by the presence of a visual face during learning, an effect referred to as “face-overshadowing” (Cook & Wilding, 1997, Cook & Wilding, 2001). Within this context, the saliency of the face interferes with the ability to attend to the voice

identity. Zäske, Mühl, & Schweinberger (2015) recently demonstrated that the face-overshadowing effect is mitigated over time. While they observed that the presence of a face initially impaired voice learning, with repeated exposure voice recognition was more robust for face-learned, compared to auditory-only learned, speakers. This is in line with the findings on face-learned in contrast to occupation-learned or name-learned voices (von Kriegstein & Giraud, 2006; von Kriegstein et al., 2008), as well as familiar voice processing (see von Kriegstein et al., 2005) where repeated day-to-day audio-visual interactions are likely typical.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was funded by a Max Planck Research Group grant to KvK.

ORCID

Corrina Maguinness  <http://orcid.org/0000-0002-6200-4109>
Katharina von Kriegstein  <http://orcid.org/0000-0001-7989-5860>

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38.
- Andrews, T., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*(3), 905–913.
- Arnold, P., & Mills, M. (2001). Memory for faces, shoes, and objects by deaf and hearing signers and hearing nonsigners. *Journal of Psycholinguistic Research*, *30*(2), 185–195.
- Arnold, P., & Murray, C. (1998). Memory for faces and objects by deaf and hearing signers and hearing nonsigners. *Journal of Psycholinguistic Research*, *27*, 481–497.
- Avidan, G., & Behrmann, M. (2014). Impairment of the face processing network in congenital prosopagnosia. *Frontiers in Bioscience*, *6*, 236–257.
- Axelrod, V., & Yovel, G. (2015). Successful decoding of famous faces in the fusiform face area. *PLoS ONE*, *10*(2), e0117126.
- Barton, J. J., & Corrow, S. L. (2016). Recognizing and identifying people: A neuropsychological review. *Cortex*, *75*, 132–150.
- Bate, S., & Bennetts, R. (2014). The rehabilitation of face recognition impairments: A critical review and future directions. *Frontiers in Human Neuroscience*, *8*, 491.
- Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. *Trends in Cognitive Sciences*, *9*(4), 180–187.
- Bellugi, U., O'Grady, L., Lillo-Martin, D., O'Grady, M., van Hoek, K., & Corina, D. (1990). Enhancement of spatial cognition in deaf children. In V. Volterra, & C. J. Erting (Eds.), *From gesture to language in hearing and deaf children* (pp. 278–298). New York: Springer-Verlag.
- Benki, J. R. (2003). Analysis of english nonsense syllable recognition in noise. *Phonetica*, *60*(2), 129–157.
- Blank, H., Anwender, A., & von Kriegstein, K. (2011). Direct Structural Connections between Voice- and Face-Recognition Areas. *The Journal of Neuroscience*, *31*(36), 12906–12915. [10.1523/JNEUROSCI.2091-11.2011](https://doi.org/10.1523/JNEUROSCI.2091-11.2011)
- 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. [doi:10.1002/hbm.22631](https://doi.org/10.1002/hbm.22631)
- Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience and Biobehavioral Reviews*, *47*, 717–734.
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*, 207–218.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.
- Bull, R., Rathborn, H., & Clifford, B. R. (1983). The voice-recognition accuracy of blind listeners. *Perception*, *12*, 223–226.
- Bülthoff, I., & Newell, F. N. (2015). Distinctive voices enhance the visual recognition of unfamiliar faces. *Cognition*, *137*, 9–21.
- Bülthoff, I., & Newell, F. N. (2017). Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception. *Visual Cognition*. [doi:10.1080/13506285.2017.1290729](https://doi.org/10.1080/13506285.2017.1290729)
- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(3), 361–380.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243–248.
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001–1010.
- Cook, S., & Wilding, J. (1997). Earwitness testimony .2. Voices, faces and context. *Applied Cognitive Psychology*, *11*(6), 527–541.
- Cook, S., & Wilding, J. (2001). Earwitness testimony: Effects of exposure and attention on the face overshadowing effect. *British Journal of Psychology*, *92*(4), 617–629.
- Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, *24*(4), 343–372.
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482–486.

- De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 27(2), 213–221.
- DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, 8, 561.
- Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, 44(2), 587–605.
- Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419–430.
- Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, 17(2), 249–261.
- Duchaine, B., & Nakayama, K. (2006a). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576–585. doi:10.1016/j.neuropsychologia.2005.07.001
- Duchaine, B. C., & Nakayama, K. (2006b). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, 16(2), 166–173.
- Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, 32(7), 827–838.
- Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, 23(5), 714–747.
- Eger, E., Schyns, P. G., & Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *NeuroImage*, 22(1), 232–242.
- 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(1), 143–156.
- Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Ewbank, M. P., & Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *NeuroImage*, 40(4), 1857–1870.
- Föcker, J., Best, A., Hölig, C., & Röder, B. (2012). The superiority in voice processing of the blind arises from neural plasticity at sensory processing stages. *Neuropsychologia*, 50, 2056–2067.
- Föcker, J., Hölig, C., Best, A., & Röder, B. (2015). Neural plasticity of voice processing: Evidence from event-related potentials in late-onset blind and sighted individuals. *Restorative Neurology and Neuroscience*, 33(1), 15–30.
- 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. doi:10.1016/j.neuropsychologia.2008.08.003
- Ghazanfar, A. A., Tureson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, 17, 425–430.
- Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: A functional magnetic resonance imaging study. *Neuropsychologia*, 47(13), 2967–2974.
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555–562.
- Grueter, M., Grueter, T., Bell, V., Horst, J., & Laskowski, W. (2007). Hereditary prosopagnosia: The first case series. *Cortex*, 43, 734–749.
- Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognise you but I can't place you. An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology*, 51A(1), 179–195.
- 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.
- Herald, S. B., Xu, X., Biederman, I., Amir, O., & Shilowich, B. E. (2014). Phonagnosia: A voice homologue to prosopagnosia. *Visual Cognition*, 22(8), 1031–1033.
- Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014a). Brain systems mediating voice identity processing in blind humans. *Human Brain Mapping*, 35(9), 4607–4619.
- Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014b). Crossmodal plasticity in the fusiform gyrus of late blind individuals during voice recognition. *NeuroImage*, 103, 374–382.
- Joassin, F., Maurage, P., & Campanella, S. (2008). Perceptual complexity of faces and voices modulates cross-modal behavioral facilitation effects. *Neuropsychological Trends*, 3, 29–44.
- Joassin, F., Pesenti, M., Maurage, P., Verreclt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex*, 47(3), 367–376.
- Jones, R. D., & Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology*, 23(3), 265–273.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302–4311.
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1476), 2109–2128.
- Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., & Horst, J. (2006). First report of prevalence of Non-Syndromic Hereditary Prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, 140A, 1617–1622.
- Kennerknecht, I., Ho, N. Y., & Wong, V. C. N. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese

- population. *American Journal of Medical Genetics Part A*, 146A (22), 2863–2870.
- Kennerknecht, I., Kischka, C., Stemper, C., Elze, T., & Stollhoff, R. (2011). Heritability of face recognition. In T. Barbu (Eds.), *Face Analysis, Modeling and Recognition Systems* (pp. 163–188). InTech.
- Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, 38(6), 618–625.
- Kress, T., & Daum, I. (2003). Developmental prosopagnosia: A review. *Behavioural Neurology*, 14(3,4), 109–121.
- von Kriegstein, K. (2012). A multisensory perspective on human auditory communication. In M. M. Murray, & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton, FL: CRC Press/Taylor & Francis.
- 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*, 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. (2006). Implicit Multisensory Associations Influence Voice Recognition. *PLoS Biology*, 4(10), e326.
- von Kriegstein, K., Kleinschmidt, A., & Giraud, A. L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex*, 16(9), 1314–1322.
- 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.
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2, 175. doi:10.3389/fpsyg.2011.00175
- Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*, 4(1), 63–74.
- Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, 61(2), 139–158.
- Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex*, 46(8), 949–964.
- Leopold, D. A., Rhodes, G., Müller, K., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences*, 272, 897–904.
- Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29(1), 117–123.
- Liu, R. R., Corrow, S. L., Pancaroglu, R., Duchaine, B., & Barton, J. (2015). The processing of voice identity in developmental prosopagnosia. *Cortex*, 71, 390–397. doi:10.1016/j.cortex.2015.07.030
- Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, 22(1), 203–211.
- Maguinness, C., & von Kriegstein, K. (2016). Visual mechanisms in the face-sensitive posterior superior temporal sulcus facilitate auditory-only speaker recognition in high levels of auditory noise. *Perception*, 45(2), (ECP Abstract Supplement), 152–152.
- McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex*, 12(1), 76–82.
- McCullough, S., & Emmorey, K. (1997). Face processing by deaf ASL signers: Evidence for expertise in distinguished local features. *Journal of Deaf Studies and Deaf Education*, 2(4), 212–222.
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56, 640–662.
- Nunn, J. A., Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: Should it be taken at face value? *Neurocase*, 7(1), 15–27.
- O'Mahony, C., & Newell, F. N. (2012). Integration of faces and voices, but not faces and names, in person recognition. *British Journal of Psychology*, 103(1), 73–82.
- Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, 68, 169–181.
- Quaranta, D., Piccininni, C., Carlesimo, G. A., Luzzi, S., Marra, C., Papagno, C., & Gainotti, G. (2016). Recognition disorders for famous faces and voices: A review of the literature and normative data of a new test battery. *Neurological Sciences*, 37(3), 345–352.
- Rivolta, D., Woolgar, A., Palermo, R., Butko, M., Schmalzl, L., & Williams, M. A. (2014). Multi-voxel pattern analysis (MVPA) reveals abnormal fMRI activity in both the “core” and “extended” face network in congenital prosopagnosia. *Frontiers in Human Neuroscience*, 8, 952.
- Roswadowitz, C., Maguinness, C., & von Kriegstein, K. (in press). Deficits in voice-identity processing: Acquired and developmental phonagnosia. In P. Belin, & S. Fruehholz (Eds.), *Oxford Handbook of Voice Perception*. Oxford: Oxford University Press.
- Roswadowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology*, 24(19), 2348–2353.
- Roswadowitz, C., Schelinski, S., & von Kriegstein, K. (2017). Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. *NeuroImage*. doi:10.1016/j.neuroimage.2017.02.064
- 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. doi:10.1016/j.neuroimage.2013.03.043
- Schall, S., & von Kriegstein, K. (2014). Functional connectivity between face-movement and speech-intelligibility areas

- during auditory-only speech perception. *PLoS ONE*, 9(1), e86325. doi:10.1371/journal.pone.0086325
- 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.
- Schiltz, C., Sorger, B., Caldara, R., Ahmed, F., Mayer, E., Goebel, R., & Rossion, B. (2006). Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. *Cerebral Cortex*, 16(4), 574–586.
- Schweinberger, S. R., Kloth, N., & Robertson, D. M. C. (2011). Hearing facial identities: Brain correlates of face-voice integration in person identification. *Cortex*, 47(9), 1026–1037.
- Sharf, D. J., & Hemeyer, T. (1972). Identification of place of consonant articulation from vowel formant transitions. *The Journal of The Acoustical Society of America*, 51(2), 652–658.
- Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception and Psychophysics*, 66(2), 352–362.
- Shilowich, B. E., & Biederman, I. (2016). An estimate of the prevalence of developmental phonagnosia. *Brain and Language*, 159, 84–91.
- Smith, H. M. J., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016a). Concordant cues in faces and voices: Testing the backup signal hypothesis. *Evolutionary Psychology*, 14(1), 1–10.
- Smith, H. M. J., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016b). Matching novel face and voice identity using static and dynamic facial images. *Attention, Perception, & Psychophysics*, 78(3), 868–879.
- Smits, R. (2000). Temporal distribution of information for human consonant recognition in VCV utterances. *Journal of Phonetics*, 28, 111–135.
- Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. *British Journal of Psychology*, 105(1), 1–16.
- Stevenage, S. V., Hugill, A. R., & Lewis, H. G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, 24(4), 409–419.
- Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology*, 23(3), 423–429.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A : Human Experimental Psychology*, 43(2), 161–204.
- Weibert, K., & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, 75, 588–596.
- Xu, X., Biederman, I., Shilowich, B. E., Herald, S. B., & Amir, O. (2015). Developmental phonagnosia: Neural correlates and a behavioral marker. *Brain and Language*, 149, 106–117.
- Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): Image or person? *Vision Research*, 49(23), 2800–2807.
- Yovel, G., & O'Toole, A. J. (2016). Recognising people in motion. *Trends in Cognitive Sciences*, 1544, 1–13.
- Zäske, R., Mühl, C., & Schweinberger, S. R. (2015). Benefits for voice learning caused by concurrent faces develop over time. *PLoS ONE*, 10(11), e0143151.
- Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, 268(1), 38–45.