Neuropsychologia 65 (2014) 1–11

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

# Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder

Stefanie Schelinski [a,*], Philipp Riedel [a,b], Katharina von Kriegstein [a,c]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany
[b] Faculty of Medicine Carl Gustav Carus at the Technische Universität Dresden, Fetscherstraße 74, 01307 Dresden, Germany
[c] Humboldt University of Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

ARTICLE INFO

*Article history:*
Received 11 April 2014
Received in revised form
25 August 2014
Accepted 18 September 2014
Available online 2 October 2014

*Keywords:*
Autism
Multisensory
Speech recognition
Voice recognition
Lip-reading
Face recognition

ABSTRACT

In auditory-only conditions, for example when we listen to someone on the phone, it is essential to fast and accurately recognize what is said (speech recognition). Previous studies have shown that speech recognition performance in auditory-only conditions is better if the speaker is known not only by voice, but also by face. Here, we tested the hypothesis that such an improvement in auditory-only speech recognition depends on the ability to lip-read. To test this we recruited a group of adults with autism spectrum disorder (ASD), a condition associated with difficulties in lip-reading, and typically developed controls. All participants were trained to identify six speakers by name and voice. Three speakers were learned by a video showing their face and three others were learned in a matched control condition without face. After training, participants performed an auditory-only speech recognition test that consisted of sentences spoken by the trained speakers. As a control condition, the test also included speaker identity recognition on the same auditory material. The results showed that, in the control group, performance in speech recognition was improved for speakers known by face in comparison to speakers learned in the matched control condition without face. The ASD group lacked such a performance benefit. For the ASD group auditory-only speech recognition was even worse for speakers known by face compared to speakers not known by face. In speaker identity recognition, the ASD group performed worse than the control group independent of whether the speakers were learned with or without face. Two additional visual experiments showed that the ASD group performed worse in lip-reading whereas face identity recognition was within the normal range. The findings support the view that auditory-only communication involves specific visual mechanisms. Further, they indicate that in ASD, speaker-specific dynamic visual information is not available to optimize auditory-only speech recognition.



## 1. Introduction

When talking to someone on the phone, recognizing what is said and recognizing who is speaking are two inherently auditory tasks. According to the conventional view, performance in these auditory tasks relies on auditory processes without contribution of visual processes (e.g. Ellis et al., 1997; Hickok and Poeppel, 2007). An alternative view proposes that visual recognition abilities are relevant also for auditory-only tasks (von Kriegstein et al., 2008). We refer to these two different views as the "auditory-only model" (Fig. 1A) and the "auditory–visual model" (Fig. 1B).

The auditory–visual model is based on behavioral findings and neuroimaging results (Blank et al., 2011; Rosenblum et al., 2007; Schall et al., 2013; Sheffert and Olson, 2004; von Kriegstein et al., 2008, 2006, 2005; von Kriegstein and Giraud, 2006). For example, several studies have shown that in auditory-only conditions (such as on the phone) typically developed individuals recognize someone's identity by voice more easily if they know this person by voice and face (Sheffert and Olson, 2004; von Kriegstein et al., 2008; von Kriegstein and Giraud, 2006). The studies showed that speaker identity recognition performance is better for voices that have been learned in a brief learning period with a voice–face video recording of a speaker in contrast to learning the voice in a matched control learning condition without the face. In the following, we refer to this improvement in behavioral performance as "face-benefit" (von Kriegstein et al., 2008). In typically developed individuals this face-benefit in speaker identity recognition is associated with enhanced

* Corresponding author. Tel.: +49 341 9940 2485; fax: +49 341 9940 2448.
*E-mail addresses:* schelinski@cbs.mpg.de (S. Schelinski), philipp.riedel@uniklinikum-dresden.de (P. Riedel), kriegstein@cbs.mpg.de (K. von Kriegstein).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.09.031
0028-3932/

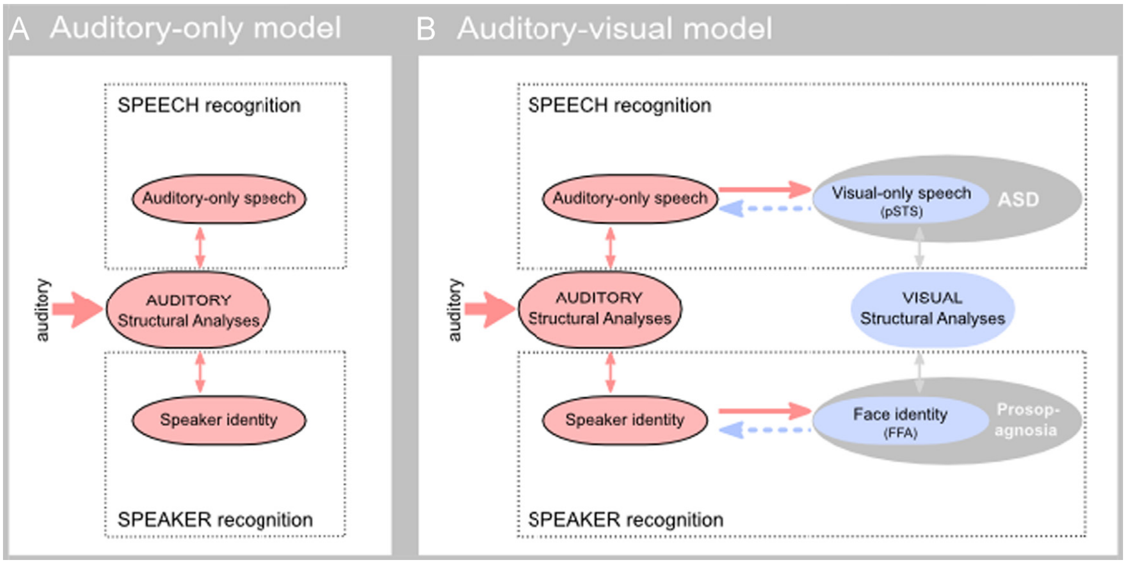


**Fig. 1.** Schematic diagram of two models for processing human auditory communication signals during auditory-only speech and speaker identity recognition. (A) Auditory-only model: In auditory-only conditions only the auditory sensory system is important for the initial sensory processing of the communication signal. (B) Auditory–visual model: The auditory and the visual system are important in auditory-only conditions. For auditory-only speech recognition, visual-only speech and auditory-only speech areas interact. Similarly for auditory-only speaker identity recognition, face and voice identity areas interact (von Kriegstein et al., 2008). Gray ellipses: Prosopagnosia is associated with a selective deficit in face identity recognition. Autism spectrum disorder (ASD) is associated with a deficit in visual-only speech recognition. pSTS = posterior superior temporal sulcus, facial movements; FFA = fusiform face area, face identity.

blood oxygen level dependent (BOLD) activity in the fusiform face area (FFA; Fig. 1B) (von Kriegstein et al., 2008), a brain region that is associated with face identity recognition (Puce et al., 1998; von Kriegstein et al., 2008).

A face-benefit also occurs for auditory-only speech recognition (Rosenblum et al., 2007; von Kriegstein et al., 2008): Typically developed individuals are better at auditory-only speech recognition for voices that have been learned in a brief period with a voice–face video recording in contrast to a matched control learning condition. In typically developed individuals this face-benefit in speech recognition is associated with enhanced BOLD activity in the posterior superior temporal sulcus (pSTS; Fig. 1B, von Kriegstein et al., 2008), a brain region that is associated with recognizing face movements (Haxby et al., 2000).

The auditory–visual model proposes that the visual face area recruitment during auditory-only speech and speaker recognition reflects a simulation process. In this view, we simulate a talking speaker's face when we hear auditory-only speech. The simulation is thought to rely on two different processes; a simulation of the face identity via the FFA and a simulation of the orofacial speech movement via pSTS. The simulation could fill in the missing visual input in auditory-only conditions and lead to the behavioral improvement, i.e. the face-benefit (von Kriegstein and Giraud, 2006; von Kriegstein et al., 2008). This view results in two hypotheses: First, a deficit in face identity recognition would lead to a lack of face-benefit in auditory-only speaker recognition. Second, a deficit in orofacial speech movement perception would lead to a lack of face-benefit in auditory-only speech recognition. Currently, there is evidence for the first hypothesis: The face-benefit in speaker identity recognition is absent in individuals with a selective face identity recognition deficit, i.e. developmental prosopagnosia (Fig. 1B, "Prosopagnosia") (von Kriegstein et al., 2008). The face-benefit for auditory-only speech recognition is normal in developmental prosopagnosics (von Kriegstein et al., 2008).

In the present study our aim is to test the second hypothesis, i.e. that a deficit in perceiving orofacial movements is associated with a lack of face-benefit in auditory-only speech recognition. To test this, we recruited a group of people with difficulties in recognizing visual-only speech (lip-reading), i.e. high-functioning autism spectrum disorder (ASD). ASD is a condition whose core features include atypical social interaction and communication (DSM-5, American Psychiatric Association, 2013; ICD-10, World Health Organization, 2004. Visual-only speech recognition is the ability to recognize speech from orofacial speech movements when only the face, but no auditory stimulus is present. Several studies have shown that individuals with ASD have difficulties with visual-only speech recognition (Gepner et al., 1996; Iarocci et al., 2010; Irwin et al., 2011; Smith and Bennetto, 2007; Williams et al., 2004; Woynaroski et al., 2013). In contrast, the ability to recognize auditory-only speech, at least under a relatively good signal-to-noise ratio, is intact (Hillier et al., 2007; Iarocci et al., 2010; Irwin et al., 2011; Smith and Bennetto, 2007; Woynaroski et al., 2013). The auditory–visual model predicts that the face-benefit in auditory-only speech recognition will be absent in individuals with difficulties in visual-only speech recognition. Therefore we expect that individuals with ASD will have difficulties in auditory-only speech recognition for speakers that they know by face, but not for those that they do not know by face (Fig. 1B, "ASD"). To test this prediction, we first evaluated the level of face processing abilities in an ASD sample with a visual-only speech recognition task (Fig. 2A) and a face identity recognition task (Fig. 2B). We expected difficulties in visual-only speech recognition (Gepner et al., 1996; Iarocci et al., 2010; Irwin et al., 2011; Smith and Bennetto, 2007; Williams et al., 2004; Woynaroski et al., 2013), and impaired but more variable performances across individuals in the ASD group in face identity recognition (Barton et al., 2004; Hedley et al., 2011). These two tests on visual face processing were independent from the main experiment. In the main experiment we tested participants' face-benefit in auditory-only speech recognition (Fig. 2C/D). For that we trained participants to identify voices and names of six speakers. In one learning condition, voices and names of three of the speakers were learned together with their face (voice–face learning). In the other learning condition, voices and names of the other three speakers were learned together with a symbol of their occupation instead of the face (voice–occupation learning). To test the face-benefit in auditory-only speech recognition we subsequently presented auditory-only speech samples from all six speakers who were learned in these two different conditions. We expected individuals with ASD to show difficulties in auditory-only