BRAIN
A JOURNAL OF NEUROLOGY

Obligatory and facultative brain regions for voice-identity recognition

Claudia Roswandowitz,[1,2] Claudia Kappes,[1] Hellmuth Obrig[1,3] and Katharina von Kriegstein[1,4,5]

Recognizing the identity of others by their voice is an important skill for social interactions. To date, it remains controversial which parts of the brain are critical structures for this skill. Based on neuroimaging findings, standard models of person-identity recognition suggest that the right temporal lobe is the hub for voice-identity recognition. Neuropsychological case studies, however, reported selective deficits of voice-identity recognition in patients predominantly with right inferior parietal lobe lesions. Here, our aim was to work towards resolving the discrepancy between neuroimaging studies and neuropsychological case studies to find out which brain structures are critical for voice-identity recognition in humans. We performed a voxel-based lesion-behaviour mapping study in a cohort of patients ($n = 58$) with unilateral focal brain lesions. The study included a comprehensive behavioural test battery on voice-identity recognition of newly learned (voice-name, voice-face association learning) and familiar voices (famous voice recognition) as well as visual (face-identity recognition) and acoustic control tests (vocal-pitch and vocal-timbre discrimination). The study also comprised clinically established tests (neuropsychological assessment, audiometry) and high-resolution structural brain images. The three key findings were: (i) a strong association between voice-identity recognition performance and right posterior/mid temporal and right inferior parietal lobe lesions; (ii) a selective association between right posterior/mid temporal lobe lesions and voice-identity recognition performance when face-identity recognition performance was factored out; and (iii) an association of right inferior parietal lobe lesions with tasks requiring the association between voices and faces but not voices and names. The results imply that the right posterior/mid temporal lobe is an obligatory structure for voice-identity recognition, while the inferior parietal lobe is only a facultative component of voice-identity recognition in situations where additional face-identity processing is required.

1 Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany
2 International Max Planck Research School on Neuroscience of Communication, Stephanstraße 1a, 04103 Leipzig, Germany
3 Clinic for Cognitive Neurology, University Hospital Leipzig, Germany
4 Humboldt University zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany
5 Technische Universität Dresden, Faculty of Psychology, Bamberger Str. 7, 01187 Dresden, Germany

Correspondence to: Claudia Roswandowitz
Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany
E-mail: roswandowitz@cbs.mpg.de

Correspondence may also be addressed to:
Katharina von Kriegstein
E-mail: katharina.von_kriegstein@tu-dresden.de

Keywords: voice-identity recognition; voxel-based lesion-behaviour mapping; posterior/mid temporal lobe; inferior parietal lobe

Abbreviations: CFMT = Cambridge Face Memory Test; MTG = middle temporal gyrus; SMG = supramarginal gyrus; STG/S = superior temporal gyrus/sulcus; VLSM = voxel-based lesion-symptom mapping

Introduction

The ability to recognize the identity of other people is a key skill for successful human communication. Impairments in person-identity recognition lead to psychosocial disabilities such as difficulties in communication, avoidance of social situations, and feelings of embarrassment and failure (Yardley *et al.*, 2008; Fine, 2012). Deficits in recognizing a person by voice can be acquired due to brain lesions or neurodegenerative diseases (Van Lancker and Canter, 1982; Neuner and Schweinberger, 2000; Hailstone *et al.*, 2011; Luzzi *et al.*, 2017a). Such deficits can also be symptoms of developmental and psychiatric disorders such as autism spectrum disorder and schizophrenia (Boucher *et al.*, 1998; Garrido *et al.*, 2009; Alba-Ferrara *et al.*, 2012; Mou *et al.*, 2013; Roswandowitz *et al.*, 2014; Schelinski *et al.*, 2016b).

Phonagnosia—a selective deficit in voice-identity processing—was first described in patients with acquired brain lesions (Assal *et al.*, 1976; Van Lancker and Canter, 1982). In these neuropsychological case studies, voice-identity recognition deficits were independent of intact face-identity recognition skills (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Neuner and Schweinberger, 2000), intact musical skills (Luzzi *et al.*, 2017b), intact language skills (Luzzi *et al.*, 2017b), and intact perceptual voice-identity analysis as assessed with voice discrimination tests (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988, 1989; Luzzi *et al.*, 2017b). The findings suggest a dedicated neural substrate for voice-identity recognition not involved in other person-recognition abilities, language skills, or even the perceptual analysis of voice-identity features (Fig. 1A). Based

on neuropsychological reports, a prime candidate area for the neural substrate for voice-identity recognition is the right inferior parietal lobe (Van Lancker *et al.*, 1988, 1989) (Fig. 1B): all but one patient (Luzzi *et al.*, 2017b) with voice-identity recognition deficits, in whom lesion information was available, had lesions in the right inferior parietal lobe.

Current neuroanatomical models of voice-identity recognition, however, do not prominently feature the inferior parietal lobe (Belin *et al.*, 2004; Blank *et al.*, 2014a). Largely based on neuroimaging findings, the temporal lobe is considered to afford voice-identity recognition (Fig. 1B). It includes the so-called temporal voice areas (TVAs) (von Kriegstein and Giraud, 2006; Campanella and Belin, 2007) located along the superior temporal gyrus/sulcus (STG/S) (Belin *et al.*, 2000; Pernet *et al.*, 2015). The TVAs have been defined based on functional MRI studies in healthy humans as well as non-human primates and dogs (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; Petkov *et al.*, 2009; Perrodin *et al.*, 2011; Andics *et al.*, 2014; Pernet *et al.*, 2015; for review see Perrodin *et al.*, 2015). Particularly the right STG/S, extending into the middle temporal gyrus (MTG), was shown to respond to identity processing of human voices, also involving left STG/S regions (von Kriegstein and Giraud, 2004; Warren *et al.*, 2006; Formisano *et al.*, 2008; Pernet *et al.*, 2015). However, evidence for a critical function of the temporal lobe in selective voice-identity recognition is scarce. To date, there is only one recent report on a selective voice-identity recognition deficit in a patient with a temporal lobe lesion (Luzzi *et al.*, 2017b). In contrast, other neuropsychological studies on patients with temporal lobe lesions report either impaired multimodal person-identity recognition (Gainotti *et al.*, 2003; Hailstone *et al.*, 2010, 2011) or
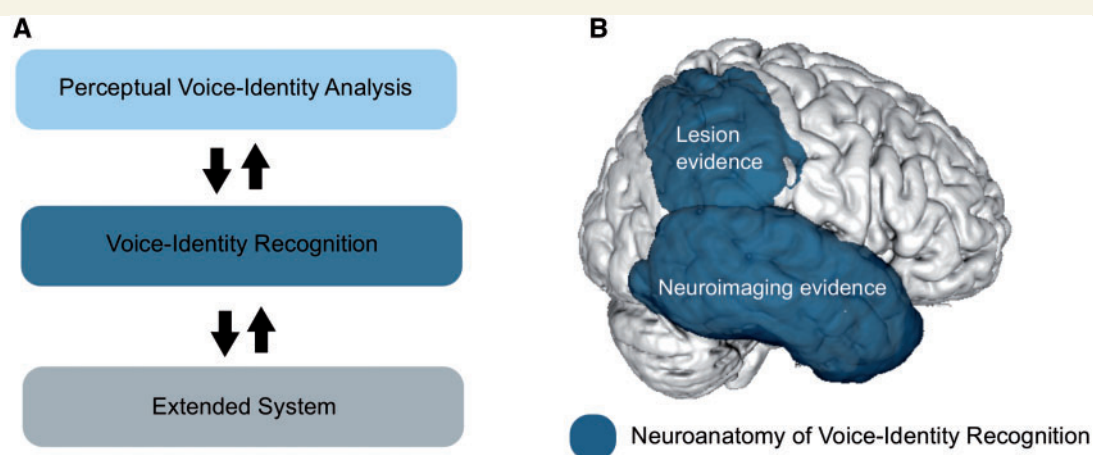

Figure 1 Schematic model on voice-identity processing and overview of potential neuroanatomical representations of voice-identity recognition. (A) Voice-identity processing is conceived as a multistage process, which comprises (i) perceptual voice-identity analysis; (ii) voice-identity recognition (feeling of familiarity for a voice); and (iii) an extended system that attributes meaning to a voice by, for instance, associating semantic information to it (Ellis *et al.*, 1997; Neuner and Schweinberger, 2000; Belin *et al.*, 2004; Blank *et al.*, 2014a; Perrodin *et al.*, 2015). **(B)** Divergent findings on the neural representation of voice-identity recognition. Most lesion studies suggest that the right inferior parietal lobe is critical for voice-identity recognition while neuroimaging studies agree that voice-identity recognition relies on the temporal lobe.

intact voice-identity recognition (Van Lancker *et al.*, 1988, 1989; Jiahui *et al.*, 2017).

Here, we addressed the apparent discrepancy between neuroimaging-based evidence for the temporal lobe to house voice-identity recognition and the fact that patient case reports consistently found an involvement of the inferior parietal lobe for voice-identity recognition tasks. We argue that although imaging studies can identify the extension of an overall network with great precision, lesion studies provide evidence that a structure is mandatory for performance of a specific task.

Previous reports on acquired phonagnosia relied either on case studies (Van Lancker *et al.*, 1988; Jiahui *et al.*, 2017; Luzzi *et al.*, 2017*b*) or group studies where lesion mapping was based on overlay representations of lesions detected on CT scans (Van Lancker *et al.*, 1988, 1989). Today, besides the advent of high spatial resolution MRI, more refined statistical analysis procedures allow for specific lesion-behaviour associations [voxel-based morphometry (VBM); Ashburner and Friston, 2000; voxel-based lesion symptom mapping (VLSM); Bates *et al.*, 2003].

Here, we used VLSM on patients with acquired unilateral brain lesions to test whether lesions in the inferior parietal and/or temporal lobe are associated with reduced voice-identity recognition. None of the patients suffered from severe language or other cognitive impairment. This was mandatory because we tested the patients on a demanding comprehensive behavioural test battery, including tests on recognizing newly-learned unfamiliar and familiar voices. To control for visual person recognition and the processing of acoustic voice features, the battery also tested face-identity recognition and vocal-pitch and vocal-timbre discrimination. All patients underwent neuropsychological assessment, pure-tone audiometry, and reported on their voice-identity recognition skills in a paper-based questionnaire. We correlated the behavioural scores with the lesion data from high resolution structural brain scans. Besides addressing our main question regarding the key structures for voice-identity recognition, the study design also allowed us to test whether recognition of voices with varying levels of familiarity involves the same or different neuroanatomical structures.

Materials and methods

Participants

We recruited 70 patients with unilateral lesions treated at the Clinic for Cognitive Neurology, University-Hospital Leipzig, Germany. Exclusion criteria were severe aphasia, severe cognitive impairment, and disorders such as dementia, schizophrenia, personality disorder, or severe depression. All patients gave written informed consent prior to testing. Data were collected in accordance with the Declaration of Helsinki and the Ethics Committee of the University of Leipzig and patients received financial compensation for their participation. Of the 70 patients, 12 were excluded from the analysis (*n* = 11

due to inability to perform all tests; *n* = 1 no neuroimaging data available). Thus, our analyses included 58 patients (31 female, 57 right-handed) (Oldfield, 1971). For an overview of the patient demographics, the results of the neuropsychological assessment, and audiometry, see Table 1. Thirty-one patients had right-hemispheric and 27 patients left-hemispheric lesions (Fig. 2A). Lesion types included ischaemic stroke (*n* = 34), intracerebral haemorrhage (*n* = 6), subarachnoid haemorrhage (*n* = 6), traumatic brain injury (*n* = 7), and extirpation of low-grade glioma (*n* = 4). Twelve of 58 patients had mild-to-moderate aphasia. Thirty-five patients had a diagnosis of mild and six patients of a moderate cognitive disorder. Detailed individual patient characteristics are provided in Supplementary Table 1.

Experimental procedures

General procedure

For computer-based tests {behavioural test battery, audiogram, attention test [Test of Attentional Performance (TAP) of neuropsychological assessment]}, participants were comfortably seated facing a 21-inch monitor displaying the visual stimuli. Auditory stimuli were presented via headphones (Sennheiser HD 280 pro). The sound level was individually adjusted to a comfortable sound pressure level. Participants' responses were recorded via a keyboard. The audiogram and the behavioural test battery were carried out in a sound-attenuated chamber at the clinic. Neuropsychological tests (except the TAP) and the questionnaire on person recognition were paper and pencil tests, and were performed in a quiet testing room. To ensure comprehension of tasks, oral and written instructions were provided prior to each test and assessment. To reduce fatigue or inattention, patients were tested in two or three sessions on separate days, depending on their preference. Patients were allowed to take breaks between blocks or sessions as required. The experimental procedure including pauses took ∼5 h. Fourteen of 58 patients participated in an older version of the test battery not including the voice-face and the face-name test (for description see below) and a different version of the questionnaire on person-identity recognition (details below).

Audiometry

We assessed hearing levels via a pure-tone (250–8000 Hz) screening audiometry (MADSEN Micromate 304, GN Otometrics).

Neuropsychological assessment

We assessed participants' performance on attention (Alertness as a subtest of TAP Version 2.2, Zimmermann and Fimm, 2009), auditory and visual-spatial working memory (Digit and Spatial Span as subtests of the WMS-R, Härting *et al.*, 2000), verbal intelligence and language comprehension (German Vocabulary Test, 'Wortschatztest' – 1st Edition, Schmidt and Metzler, 1992), and on associative learning abilities ('Face-name learning', GNL, Schuri and Benz, 2000) (Supplementary material).

Questionnaire on person-identity recognition

We developed a paper-based questionnaire to assess subjective abilities in everyday person-identity recognition for the time

Table 1 Demographic details, neuropsychological measures, and behavioural measures

	n	Mean	Median	SD	Range
Demographic details					
Age	58	47.95	50.00	11.59	46.00
Time since onset (month)	57	46.02	30.00	51.51	271.00
Education (years)	58	10.50	10.00	1.23	4.00
Hearing level (dB)	58	18.11	16.25	8.45	35.84
Neuropsychological tests					
Digit span (PR)	58	32.15	29.00	22.16	88.00
Spatial span (PR)	58	39.35	30.00	24.95	90.50
TAP (alertness) (PR)	54	32.77	30.00	23.64	85.00
German Vocabulary Test (PR)	54	54.51	54.00	20.61	83.80
Face-name learning	53				
Cued + sum (PR)		61.34	66.50	34.36	100.00
Memory (%)		87.56	100.00	22.04	117.00
Behavioural test battery					
Newly-learned voice recognition					
Newly-learned voice tests (%)	43	59.97	58.34	10.30	41.22
Voice-name test (%)	58	56.14	53.34	12.91	58.33
Voice-face test (%)	43	64.08	62.00	11.66	48.00
Familiar voice recognition					
Famous voice test (z)	57	0.00	0.07	0.70	3.03
Familiarity decision (d′)	57	1.11	1.23	0.84	5.32
Semantic association (%)	57	69.60	72.00	18.81	75.00
Acoustical control tests					
Vocal-pitch test (cent)	58	125.17	125.20	66.05	231.68
Vocal-timbre test (SER)	55	9.73	10.80	4.62	16.15
Visual control tests					
CFMT (%)	57	66.79	69.44	15.52	58.34
Face-name test (%)	43	60.54	62.67	19.15	76.00

The table displays mean scores, median scores, standard deviations (SD), range of scores on demographical, neuropsychological measures, and on each behavioural test respectively for all patients who completed the test (n). Hearing levels were averaged over both ears. The newly-learned voice test score is a composite score of the voice-name and voice-face test. The famous voice test score is a composite score of the familiarity decision and semantic association subscores. PR = percentage rank; SER = spatial envelope ratio; TAP = Test of Attentional Performance.

before and after lesion onset. Participants rated their general person-, voice-, and face-identity recognition abilities on a scale from 1 (excellent) to 5 (very poor). Further, we asked participants about the cues they use for person-identity recognition (e.g. person as a whole, face, voice, clothing, or posture). An English version of the questionnaire is available online (http://kriegstein.cbs.mpg.de/questionnaire/questionnaire_patients.pdf). The older version of the questionnaire assessed participants' abilities only for the time after lesion onset.

Behavioural test battery

The behavioural test battery was used in a previous study on healthy participants (Roswandowitz et al., 2014). Here, we conducted the same version except for an adaption of the famous voice test (see below and Supplementary material).

Newly-learned voice-identity recognition: voice-name and voice-face test

In the voice-name and voice-face test, participants learned and recognized six previously unfamiliar voices per test. Unfamiliar voices were learned either together with a written first name (voice-name test) or with the picture of the speaker's face (voice-face test).

Stimuli and presentation software: The auditory stimuli were recorded from 12 native speakers of standard German (six female, age range 21–32 years). The visual stimuli comprised photographic images of the speakers' faces. For more details on the stimuli, see the Supplementary material. The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure and analysis: The procedure of the voice-name and the voice-face test was the same, although different stimuli and different speakers were used in both tests. The tests contained a 'female-voice' part and a 'male-voice' part, identical in structure. The parts were structured into four learning and five testing phases. Phases were presented in alternating order, except that there was no learning phase before the fifth testing phase. During learning, participants heard declarative sentences (each ∼2 s) spoken by three speakers. Each sentence presentation was accompanied either by a name (voice-name test) or a face (voice-face test). In the respective tests, participants had to learn these voice-name or voice-face associations. During the testing phases, participants listened to different auditory sentences and performed a three-alternative forced choice task in which they selected either the name
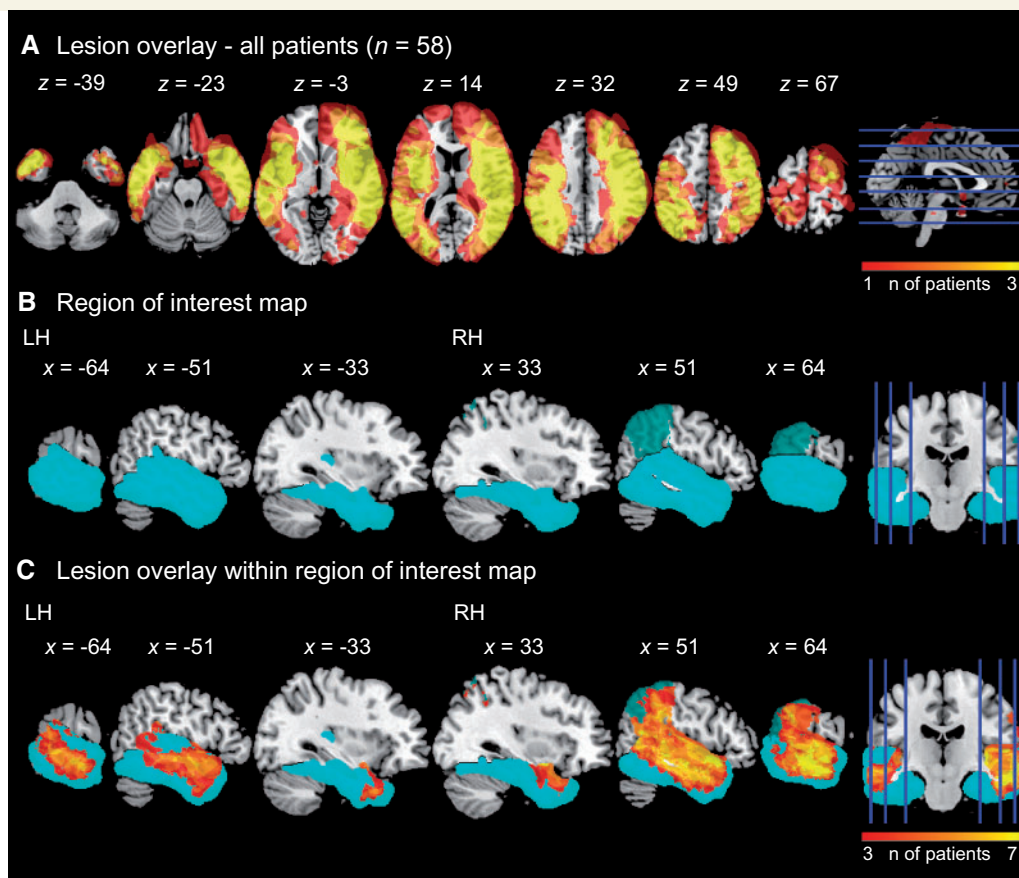
Figure 2 Lesion overlay and region of interest map. (**A**) Lesion overlay map of all patients over the whole brain. Left hemisphere is left on coronal slice. The heatmap displays the number of patients with lesions in that respective area. Coordinates refer to MNI space. (**B**) Region of interest map. Anatomical region of interest map covered the bilateral temporal lobe (cyan) and the right inferior parietal lobe (turquoise). (**C**) Lesion overlay map of all patients within the region of interest map. Areas covered by the heatmaps are lesioned in at least three patients.

(voice-name test) or the face (voice-face test) that matched the voice. To avoid prosody-driven identity recognition, we presented different types of sentences in each testing phase (2 s declarative, 2 s interrogative, and 0.7 s declarative). Each test— voice-name and voice-face—took ~20 min. For details see Supplementary material.

As a measure of newly-learned voice-identity recognition performance, we calculated the percent correct score for each participant for all trials over all five testing phases of the female and male parts, for the voice-name and voice-face test separately.

Familiar voice-identity recognition: famous voice test

In the famous voice test, participants listened to samples of famous and non-famous voices and categorized them as belonging to either familiar or unfamiliar people ('familiarity decision'). If they categorized a voice as familiar, participants were asked to associate semantic information with the speaker's voice ('semantic association').

Stimuli and presentation software: The auditory stimulus set contained voice samples of famous (n = 42) (Supplementary material) and non-famous (n = 20) German speakers. Each sample lasted 5 s. Based on a pilot study, only voices with high familiarity ratings were included in the stimulus set

(Supplementary material). The test was implemented in Presentation software (Neurobehavioural Systems, Inc.) and responses were recorded via keyboard.

Procedure and analysis: During the test, famous and non-famous voice samples were presented randomly intermixed. After each sound sample, participants were asked to categorize the voices as familiar or unfamiliar ('familiarity decision'). If on a given trial the voice sample was categorized as unfamiliar, the next voice sample was presented. If participants categorized the voice as familiar, irrespective of whether the voice was famous or non-famous, afterwards they performed a four-alternative forced choice task ('semantic association'). A choice of three celebrities (name and face picture) and a question mark were presented on the screen (Supplementary material).

After the test, we conducted a computer-based questionnaire assessing each participant's familiarity with the presented celebrities. We used this for the analysis, where we individually reclassified all famous voices (n = 42) into subjectively familiar and non-familiar voice categories (Supplementary material). We analysed the data by applying detection theory (Macmillan and Creelman, 2004). We computed indices of sensitivity (d-prime, d') of voice familiarity decision assuming the yes-no decision model [$d' = z(H) - z(FA)$]. Each famous voice subjectively classified as familiar and correctly classified

as familiar was considered a 'hit' (H) and each famous voice subjectively classified as unfamiliar incorrectly classified as familiar was considered a 'false alarm' (FA).

Additionally, we computed the proportion of those familiar voices that were correctly matched with the name and face of the respective celebrity ('semantic association') and those that were not correctly matched or for which the question mark was chosen. The complete test took ~45 min with a break between the famous voice test (30 min) and the follow-up survey on celebrity familiarity (15 min).

Acoustic control tests: vocal-pitch and vocal-timbre test

We measured the individual just noticeable differences (JND) for vocal pitch (vocal-pitch test) and vocal timbre (vocal-timbre test) in vowel sounds by using an adaptive tracking procedure. Pitch and timbre are important acoustic voice features for voice-identity recognition (Lavner et al., 2000; Gaudrain et al., 2009). Each test took ~15 min (Supplementary material).

Visual control tests: Cambridge Face Memory Test and face-name test

With the Cambridge Face Memory Test (CFMT) (Duchaine and Nakayama, 2006), we tested the ability to recognize newly-learned faces. Participants first memorized six male faces. They were then asked to recognize one of three presented faces (one learned, two unfamiliar). The CFMT comprises three different test sections: same images, novel images, and novel images with added Gaussian noise. There was a total of 72 items. The test took ~15 min.

In the face-name test (Roswandowitz et al., 2014), participants learned and recognized previously unfamiliar faces. Six male faces were associated with a written first name. After learning, participants were presented with a novel picture of one of the faces and were asked to select the corresponding name from six alternatives. The test took ~30 min (Supplementary material).

Imaging methods

For 56 patients structural high-resolution MRI scans and for two patients CT scans were available. MRI scans were acquired on a 3 T Siemens MRI system (Siemens Trio® or Verio® system, Siemens Medical Systems) including 3D T_1-weighted (1 mm^3 isotropic voxels) and FLAIR images. Lesions were manually delineated in all three planes (axial, coronal, sagittal) on each slice of the T_1 images using MRIcron (Rorden and Brett, 2000). The FLAIR images served as a reference. Lesion delineation was performed by an experienced neurologist (H.O.) who was blind to the individual patient's performance in the behavioural test battery. The T_1 images were transformed into standard stereotactic space (MNI) using SPM8 (www.fil.ion.ucl.ac.uk/spm). The unified segmentation approach was applied (Ashburner and Friston, 2005) and the estimation of normalization parameters was restricted to healthy tissue using the predefined lesion mask (cost function masking, Brett et al., 2001). The CT images (3–5 mm slice thickness and ⩽0.5 mm in-plane resolution) were analysed by the same normalization package ('clinical toolbox'; nitrc.org/projects/clinicaltbx/). These images were resampled to yield the same 1 mm^3 isovoxel resolution as the MRI images.

Voxel-based lesion-behaviour analysis

To identify systematic relationships between lesioned brain regions and behavioural measures, we performed VLSM analyses (Bates et al., 2003) as implemented in VLSM 2.55 (https://langneurosci.mc.vanderbilt.edu/resources.html) in a Matlab environment (version 8.2, The MathWorks, Inc., MA, USA). For the VLSM analyses, we included only voxels in which at least three patients had a lesion (Fig. 2A). For each behavioural test, the spatially normalized lesion maps and corresponding behavioural scores were used to create t-maps. On a voxel-by-voxel basis, the VLSM algorithm runs a general linear model comparing behavioural performances in patients with a lesion versus patients without a lesion in the respective voxel. Thus, lesion status of the voxel (0 = intact, 1 = lesioned) is the independent variable and behavioural measures the dependent variables. We applied general linear models with behavioural measures of each test separately and also with composite scores. To compute composite scores of measures with different units (e.g. per cent correct scores of unfamiliar voice tests and d′ scores of the famous voice test), we z-transformed them. We also calculated general linear models with covariates of no interest. Such a procedure allows the analysis of variance in the predictor variable while minimizing variance induced by covariates of no interest (Baldo et al., 2013; Rogalsky et al., 2015; Binder et al., 2016). Patients with traumatic brain injury often have diffuse axonal injury that is associated with cognitive impairments and is difficult to precisely detect with the MRI sequences that we used (for review, see Sharp et al., 2014). We therefore conducted a supplementary analysis excluding the seven patients with traumatic brain injury.

Significance threshold

Statistical maps were thresholded at voxel-wise $P < 0.01$. They were then corrected for multiple comparisons based on cluster size with respect to 1000 permutations in which behavioural scores were randomly reassigned (Kimberg et al., 2007; Wilson et al., 2010). After permutation, clusters with a corrected $P < 0.05$ were considered significant. We applied the same significance threshold for whole-brain and region of interest analyses. For the behavioural analyses, we applied a significance threshold of $P < 0.05$. We used Pearson correlation to compute correlations and ANOVAs or Mann-Whitney U-tests to assess group differences as implemented in IBM SPSS Statistics for Windows, Version 24 (Armonk, NY: IBM Corp.). When group variances were equal (tested with Levene's test), we used ANOVAs, and if not, we applied Mann-Whitney U-tests. For significant lesion-behaviour associations and behavioural group differences, effect sizes d were calculated using the following equation based on the T-statistics (t) and the number of participants (n) $d = 2t / \sqrt{n}$.

Region of interest definition

We created a region of interest map including bilateral temporal lobes and right inferior parietal lobe (Fig. 2B) to investigate brain structures associated with voice-identity recognition. The region of interest was based on atlases provided in FSL (Smith et al., 2004, http://www.fmrib.ox.ac.uk/fsl/fslview). We extracted probabilistic maps of the bilateral temporal lobe from the MNI Structural Atlas (Mazziotta et al., 2001) and of the right inferior parietal lobe from the

Jülich Histological Atlas (Caspers *et al.*, 2013). The inferior parietal lobe map contained the areas PF, PFm, and Pga. The resulting map covered parietal regions previously associated with voice-identity recognition deficits including supramarginal gyrus (SMG) and angular gyrus (Van Lancker *et al.*, 1988). We extracted the maps from different atlases, because (i) the parietal lobe map of the MNI Structural Atlas does not contain definitions of substructures and also includes regions not reported to be critically involved in voice-identity recognition (i.e. lateral occipital cortex, postcentral gyrus, postcentral gyrus); and (ii) a temporal lobe map is not provided in the Jülich Histological Atlas. Based on visual inspection, we chose a threshold of 10% to restrict maps to anatomically meaningful brain regions. The final region of interest map was a union of the bilateral temporal lobe and the right inferior parietal lobe map.

VLSM results reporting

We report MNI coordinates for the clusters' centre of mass. For anatomical labelling we used three atlases provided by the FSL Anatomy Toolbox: (i) the Jülich Histological Atlas (Eickhoff *et al.*, 2005) for subclassification of the Heschl's gyrus; (ii) the MNI Structural Atlas (Mazziotta *et al.*, 2001) to classify the brain lobes; and (iii) the Harvard-Oxford Cortical and Subcortical Structural Atlas (Desikan *et al.*, 2006) for all other structures.

Results

Behavioural results

In all voice- and face-identity recognition tests, patients performed significantly above chance level. Patients' average performance in the neuropsychological assessment ranged between the 33rd and 61st percentile but always above the 2nd percentile—the cut-off for impaired performance (Wechsler, 1997). All test results including the JNDs for vocal pitch and timbre are reported in Table 1. Test performance was not significantly influenced by demographic variables (Supplementary material). For patients' subjective abilities on person-identity recognition see Supplementary material and Supplementary Tables 2–4.

Since the comparison of patients with right- versus left-hemispheric lesions has a long tradition in voice-identity recognition research (Assal *et al.*, 1981; Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1989; Lang *et al.*, 2009), we first compared voice-identity recognition performances in these two groups. The familiar voice-identity recognition performance was comparable for the groups with left- and right-hemispheric lesions (composite score of familiarity decision and semantic association subscores of the famous voice test: $Z = 1.43$, $P = 0.237$, $d = 0.38$). In contrast, patients with right- compared to left-hemispheric lesions performed worse for newly-learned voice-identity recognition (composite score of voice-face and voice-name test: $Z = -2.29$, $P = 0.022$, $d = 0.70$). For completeness, we report group differences for the other behavioural measures in Supplementary Table 5.

For information, we report performances on the voice-name test of an age-matched control group without known brain damage (Roswandowitz *et al.*, 2014) in Supplementary Table 6.

VLSM results

Lesion coverage

We checked whether the lesion overlay for all 58 patients (Fig. 2A) covered previously reported voice-identity sensitive areas with lesions in ≥3 patients. These areas included the temporal lobe and inferior parietal lobe (Fig. 1B). Within our anatomical temporal lobe map, all *a priori* defined voice-sensitive regions, which have been reported in neuroimaging studies (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003, 2005; von Kriegstein and Giraud, 2004; Warren *et al.*, 2006; Blank *et al.*, 2011, 2014*a*; Chandrasekaran *et al.*, 2011; Bonte *et al.*, 2014; Roswandowitz *et al.*, 2017), were covered by lesions (≥3 patients) (Fig. 2B and C). Within the right inferior parietal lobe map, all areas with *a priori* anatomical hypotheses were covered by lesions (≥3 patients) (Fig. 2B and C). For an overview of temporal lobe and inferior parietal lobe structures not covered by lesions in ≥3 patients see Supplementary Table 7.

Voice-identity recognition

First, we investigated which lesion locations were associated with decreased voice-identity recognition performance irrespective of voice familiarity. We performed a VLSM analysis including the composite score of the z-transformed performance in newly-learned (voice-name and voice-face test) and familiar (familiarity decision and semantic association of the famous voice test) voice scores as the dependent variable. There was a statistical brain-behaviour association in a large right posterior/mid temporal lobe cluster with its centre of mass in the planum temporale extending into the mid/posterior STG and inferior parietal lobe with its centre in the SMG (Fig. 3A and Table 2). This cluster remained significant even when we controlled for processing of acoustic voice features by entering vocal-pitch and vocal-timbre JNDs as covariates of no interest into the analysis ($x = 55$, $y = -27$, $z = 14$, $T = 4.43$, $P = 0.001$, effect size = 1.46, lesion volume = 19 532). We calculated an additional analysis with hearing level (mean over both ears) and lesion volume as covariates of no interest since overall lesion volume and lesser hearing capacities may have a non-specific effect on test performances. This analysis disclosed a right posterior/mid temporal lobe cluster with its centre in the posterior STG (Fig. 3A and Table 2). Notably, this cluster again extended to the right inferior parietal lobe. At the whole-brain level, there was no further lesion cluster significant.
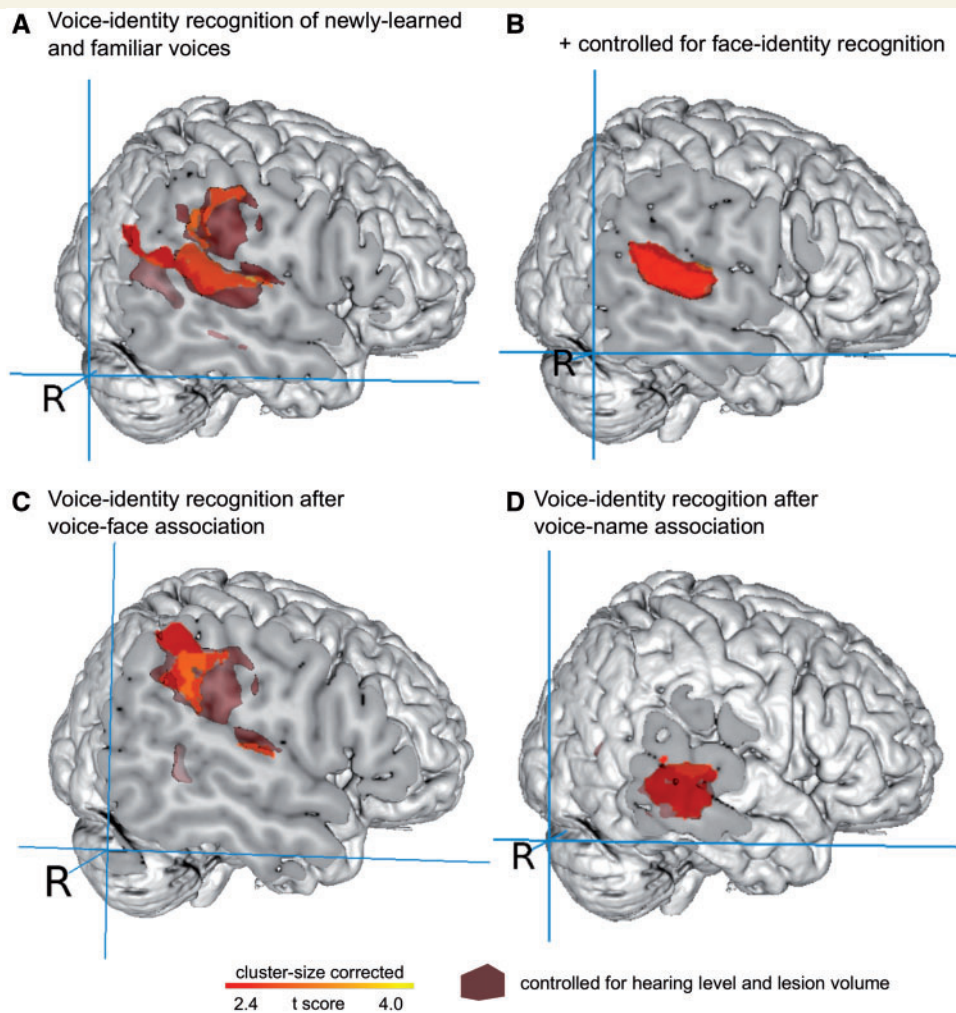
Figure 3 Lesions in the temporal and inferior parietal lobe associated with difficulties in voice-identity recognition. (**A**) VLSM results of the composite score of the newly-learned and familiar voice score (*n* = 42). (**B**) VLSM results for the same score as shown in **A**, but controlled for performance in face-identity recognition (composite score of CFMT and face-name test, *n* = 40). (**C**) VLSM results of the voice-face test (*n* = 43). (**D**) VLSM results of the voice-name test (*n* = 58). VLSM results that are controlled for hearing level and lesion size are overlaid as dark red surfaces on each image. The analyses were restricted to the bilateral temporal lobe and right inferior parietal lobe. For visualization, voxels are shown at *P* < 0.01 uncorrected.

Voice-identity recognition controlled for face-identity recognition

Next, we tested whether the statistical brain-behaviour association for voice-identity recognition in the right posterior/mid temporal and inferior parietal lobe was specific for voice-identity independent of face-identity recognition abilities. This is important because a concomitant impairment of voice- and face-identity recognition would suggest a supramodal person-identity recognition region in contrast to a region specifically critical for voice-identity recognition. We entered the z-transformed composite score of the CFMT and face-name test as a covariate into the VLSM analysis reported in the previous paragraph (i.e. with the composite score of newly-learned and familiar voice score as dependent variable). There was a significant brain-behaviour association in the right posterior/mid temporal lobe (Fig. 3B) with the centre of mass in the planum

temporale extending into mid/posterior STG (Table 2). After adding hearing level and lesion volume as covariates, the same right temporal lobe cluster remained significant. There was no significant cluster at a whole-brain level.

The results so far suggested that both the posterior/mid temporal and inferior parietal lobe are critical for voice-identity recognition, but that the posterior/mid temporal lobe might be specifically involved in voice-identity recognition, while the inferior parietal lobe might rather afford voice-identity recognition when it involves face-identity aspects.

Voice-identity recognition after voice-face and voice-name association

To explicitly assess the role of face information during voice-identity recognition in the inferior parietal lobe, we separately looked at brain-behaviour associations for

Table 2 Overview of lesion-cluster coordinates associated with voice-identity recognition ($P < 0.05$ cluster-size corrected, 1000 permutations)

Region	x	y	z	Max T	P	Effect size d	Voxels, n	x	y	z	Max T	P	Effect size d	Voxels, n
	Voice-identity recognition (newly-learned and familiar voices)							**+ Hearing level and lesion volume**						
Right temporal lobe														
Planum temporale	57	−34	16	3.96	0.005	1.22	12 460							
Posterior STG								56	−37	13	3.51	0.004	1.08	15 656
	Voice-identity recognition Face-identity recognition controlled							**+ Hearing level and lesion volume**						
Right temporal lobe														
Planum temporale	59	−30	12	3.38	0.038	1.07	5442	59	−30	12	3.26	0.038	1.03	5442
	Voice-face test							**+ Hearing level and lesion volume**						
Right parietal lobe														
Parietal operculum cortex	59	−33	32	3.36	0.003	1.03	6921	57	−34	28	3.30	0.013	1.00	8499
	Voice-name test							**+ Hearing level and lesion volume**						
Right temporal lobe														
Posterior MTG	62	−31	0	3.03	0.030	0.80	3532	60	−35	−1	3.73	0.023	0.98	8176

Centre of mass coordinates are reported in MNI space (in mm). Anatomical labels according to the Harvard-Oxford Cortical and Subcortical Structural Atlas (Desikan *et al.*, 2006) provided by the FSL Anatomy Toolbox (Smith *et al.*, 2004). Voxels, n = number of voxels; x, y, z = MNI coordinates.

identity recognition of voices that were associated with faces (voice-face test) and of voices associated with names (voice-name test). For the voice-face test, a lesion cluster associated with decreased performance was located in the right inferior parietal lobe with the centre of mass in the parietal operculum (Fig. 3C and Table 2) extending into the posterior/mid temporal lobe. This cluster remained significant after adding hearing level and lesion volume as covariates (Fig. 3C and Table 2). For the voice-name test, the lesion cluster was located in the right posterior/mid temporal lobe with the centre of mass in the posterior MTG (Fig. 3D and Table 2) and remained significant after adding hearing level and lesion volume as covariates. Note that although the significant peak clusters of the voice-face and voice-name tests were located in distinct lobes, there was an overlap of lesions in the right posterior temporal lobe (posterior STG) when controlling for hearing level and lesion volume (Fig. 3C and D).

Newly-learned and familiar voice-identity recognition

Neuroimaging studies have shown that processing voice identity with varying levels of familiarity (unfamiliar, newly-learned, and familiar voices) involves partly distinct brain regions (von Kriegstein and Giraud, 2004; Birkett *et al.*, 2007; Bethmann *et al.*, 2012; for a review see Maguinness *et al.*, submitted for publication). We here tested whether there is an association between different lesion clusters and recognition of newly-learned and familiar voices.

To assess lesion clusters associated with the recognition of newly-learned voices, we performed a VLSM analysis including the newly-learned voice score as dependent variable. We found a significant brain-behaviour association in

a right inferior parietal lobe cluster with the centre of mass in the posterior SMG and in four right posterior/mid temporal lobe clusters. The centres of mass of these clusters were located in the posterior STG, posterior MTG, planum temporale, and Heschl's gyrus. Lesion clusters in the right posterior/mid temporal (posterior MTG) and right inferior parietal lobe (posterior SMG) remained significant after controlling for hearing level and lesion volume (Fig. 4A and Table 3).

For familiar voice-identity recognition (familiar voice score as dependent variable), there was a significant brain-behaviour association in the left posterior temporal lobe with the centre of mass in the posterior MTG (Fig. 4B and Table 3). However, after adding hearing level and lesion volume as covariates, no cluster remained significant, indicating that the association needs to be interpreted with caution. Further analyses with the separate measurements of the famous voice test (i.e. familiarity decision and semantic association) revealed no significant results (for a report at a more lenient threshold see Supplementary Fig. 1 and Supplementary Table 8).

The supplementary analysis excluding patients with traumatic brain injury replicated the findings of the initial analysis including all patients. Only the association between the left posterior temporal lobe and familiar voice-identity recognition became weaker (non-significant at $P < 0.01$ uncorrected).

Discussion

We here assessed the contributions of temporal and inferior parietal lobe structures to voice-identity recognition by voxel-based lesion-behaviour mapping. The study included
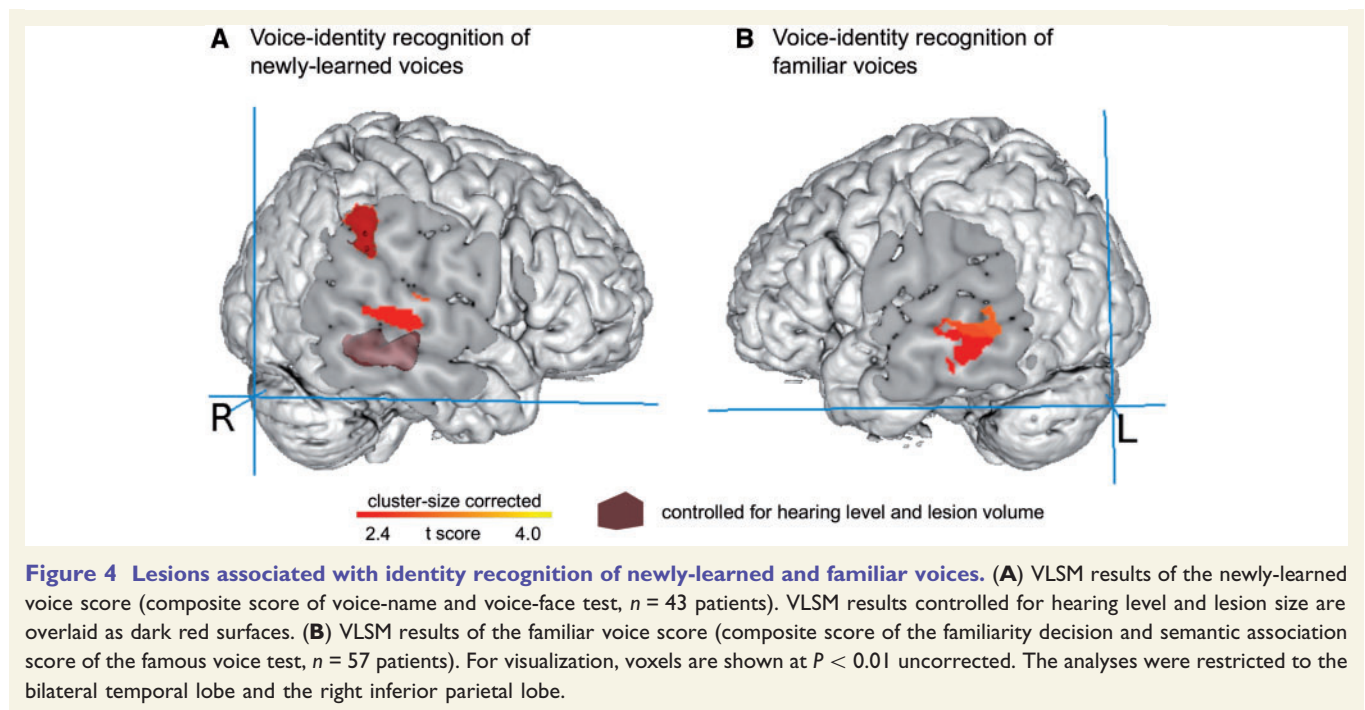
Figure 4 Lesions associated with identity recognition of newly-learned and familiar voices. (**A**) VLSM results of the newly-learned voice score (composite score of voice-name and voice-face test, *n* = 43 patients). VLSM results controlled for hearing level and lesion size are overlaid as dark red surfaces. (**B**) VLSM results of the familiar voice score (composite score of the familiarity decision and semantic association score of the famous voice test, *n* = 57 patients). For visualization, voxels are shown at $P < 0.01$ uncorrected. The analyses were restricted to the bilateral temporal lobe and the right inferior parietal lobe.
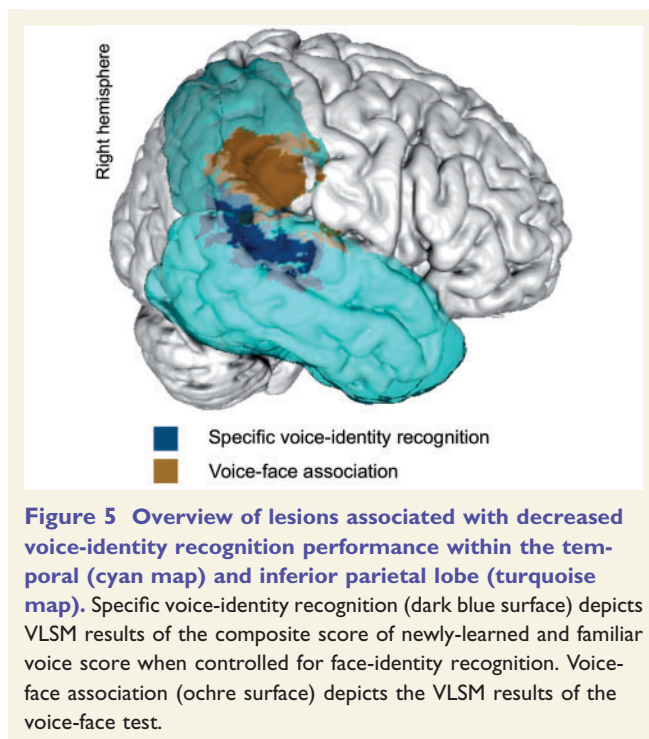
Table 3 Overview of lesion-cluster coordinates associated with newly-learned and familiar voice-identity recognition ($P < 0.05$ cluster-size corrected, 1000 permutations)

Region	x	y	z	Max T	P	Effect size, d	Voxels, n	x	y	z	Max T	P	Effect size, d	Voxels, n
	Newly-learned voice-identity recognition (voice-name and voice-face test)							**+ Hearing level and lesion volume**						
Right parietal lobe														
Posterior SMG	58	−37	32	3.25	0.002	1.00	1971	57	−39	36	3.02	0.031	0.92	2133
Right temporal lobe														
Posterior STG	62	−27	3	2.69	0.003	0.82	1116							
Posterior MTG	48	−37	−3	2.62	0.018	0.80	90	58	−31	−5	3.24	0.019	0.99	4087
Planum temporale	59	−17	8	2.77	0.023	0.85	59							
Heschl's gyrus (Te 1.1, Te 1.0)	44	−23	10	2.50	0.036	0.76	17							
	Familiar voice-identity recognition (familiarity and semantic association)							**+ Hearing level and lesion volume**						
Left temporal lobe														
Posterior MTG	−55	−33	−8	2.99	0.04	0.79	4447	–	–	–	–	–	–	–

Centre of mass coordinates are reported in MNI space (in mm). Anatomical labels according to the Harvard-Oxford Cortical and Subcortical Structural Atlas (Desikan *et al.*, 2006) provided by the FSL Anatomy Toolbox (Smith *et al.*, 2004). Voxels, *n* = number of voxels; *x, y, z* = MNI coordinates.

an extensive behavioural test battery and high-resolution brain imaging in a cohort of patients with acquired brain lesions. We found that (i) lesions in both the right posterior/mid temporal and right inferior parietal lobe were associated with poorer performance on voice-identity recognition in contrast to when these regions were intact; (ii) when face-identity recognition abilities were factored out, only lesions in the right posterior/mid temporal lobe were associated with poorer voice-identity recognition abilities; (iii) right inferior parietal lobe lesions were associated with poorer voice-identity recognition abilities only when

voices were paired with faces, but not when paired with names. These main findings are summarized in Fig. 5. The findings provide an important step towards resolving the current discrepancy between neuroimaging and neuropsychological case reports. They show that the right posterior/mid temporal lobe is critical for successful recognition of voice identities. Furthermore, the findings qualify the contribution of the right inferior parietal lobe to voice-identity recognition and provide a potential explanation for its critical involvement in voice-identity recognition in previous case reports (Van Lancker *et al.*, 1988, 1989). In addition

Figure 5 Overview of lesions associated with decreased voice-identity recognition performance within the temporal (cyan map) and inferior parietal lobe (turquoise map). Specific voice-identity recognition (dark blue surface) depicts VLSM results of the composite score of newly-learned and familiar voice score when controlled for face-identity recognition. Voice-face association (ochre surface) depicts the VLSM results of the voice-face test.

to the key findings, our results demonstrate that identity recognition of newly-learned and famous voices is supported by partly dissociated brain regions.

Temporal lobe: voice-identity recognition

Our findings support a central assumption of current voice-identity processing models (Belin *et al.*, 2004; Blank *et al.*, 2014*b*; Perrodin *et al.*, 2015; Roswandowitz *et al.*, In press): the posterior/mid temporal lobe is critical for voice-identity recognition and this process is relatively independent from acoustic voice feature, face-identity, and speech processing. However, hitherto, causal evidence for the involvement of the temporal lobe in such specific voice-identity recognition has been largely lacking (Luzzi *et al.*, 2017*b*): case studies on patients with lesions in temporal lobes reported either unimpaired voice-identity recognition (Van Lancker *et al.*, 1988, 1989; Jiahui *et al.*, 2017) or, particularly with lesions in the anterior temporal lobe, a multimodal person-recognition deficit, affecting also face and name recognition (Gainotti *et al.*, 2003, 2008; Hailstone *et al.*, 2010, 2011; Luzzi *et al.*, 2017*a*). A multimodal person-recognition deficit associated with temporal lobe lesions was also found in a patient group study with a similar methodological approach to ours (Hailstone *et al.*, 2011). In contrast, in the present study, the brain-behaviour association in right posterior/mid temporal lobe structures for voice-identity recognition was independent from abilities for face-identity recognition and present in patients with relatively intact language skills.

It is an open question why many previous reports on patients with temporal lobe lesions found no association with voice-identity recognition difficulties, but only difficulties with the discrimination of two voices (Van Lancker *et al.*, 1988, 1989) or no difficulties at all for voice-identity processing (Jiahui *et al.*, 2017). One possibility is that the cases did not have lesions in subregions of the temporal lobe that are relevant for voice-identity recognition. Furthermore, a VLSM analysis might have more statistical power to detect associations between the location of brain injury and impaired voice-identity recognition because of the voxel-wise statistics on continuous behavioural data.

Inferior parietal lobe: voice-face representations

Our results showed that the supramarginal gyrus of the right inferior parietal lobe was predominantly involved in voice-identity recognition when a voice was paired with a face. This finding may explain why previous lesion studies reported right inferior parietal lobe lesion association with voice-identity recognition impairments (Van Lancker *et al.*, 1988, 1989). The tasks used in these studies required patients to match a famous voice to a display of faces (and their corresponding names). Such tasks are not usually used in neuroimaging studies on voice-identity recognition, which likely explains the paucity of parietal lobe responses in neuroimaging studies on voice processing (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; Andics *et al.*, 2010). However, the results of two neuroimaging studies, which investigated crossmodal voice-face priming and voice-face learning, respectively, are congruent with the suggestion that the right inferior parietal lobe is involved in a representation of person-related voice and face information (von Kriegstein and Giraud, 2006; Holig *et al.*, 2017). They showed that the inferior parietal lobe is involved in voice-identity recognition for voices learned with faces, but not with names (von Kriegstein and Giraud, 2006). In the same vein, Holig *et al.* (2017) showed that voices primed by a face of the same speaker as compared to another speaker's face elicited higher brain responses in the right inferior parietal lobe (i.e. angular gyrus), again suggesting a particular role in the association of voices and faces.

The role of the posterior/mid STG for voice-identity recognition

The lesion sites in the present study that correlated with difficulties in selective voice-identity recognition (i.e. controlled for low-level acoustic and face processing) were located in the right posterior/mid STG/S and right planum temporale. This is in agreement with findings in developmental disorders with voice-identity recognition deficits including high functioning autism spectrum disorder and developmental phonagnosia (Schelinski *et al.*, 2016*a*; Roswandowitz *et al.*, 2017). In these studies the right posterior STG/S region was dysfunctional during voice identification of newly-learned voices in autistic populations

(Schelinski *et al.*, 2016*a*) and the right planum temporale was dysfunctional on the same task in a case with developmental apperceptive phonagnosia (Roswandowitz *et al.*, 2017). Apperceptive phonagnosia refers to cases with difficulties in voice-identity recognition that are also associated with voice-discrimination impairments (Hailstone *et al.*, 2011; Roswandowitz *et al.*, 2014).

We found no right anterior temporal lobe lesion cluster associated with voice-identity recognition task performance. This contrasts with a recently reported patient who had lesions in the right anterior temporal lobe (as well as lenticular and caudate nuclei) and impaired voice-identity recognition (Luzzi *et al.*, 2017*b*) and neuroimaging studies that found right anterior temporal lobe region responses for voice-identity recognition (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; Andics *et al.*, 2010). Why we found no lesion-behaviour association in the right anterior temporal lobe with voice-identity recognition remains an open question because the lateral region of the anterior temporal lobe, which has been implicated in voice-identity recognition (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; Andics *et al.*, 2010; Luzzi *et al.*, 2017*b*), was covered by lesions in at least three patients in our study. Also, our test battery included voice-identity recognition tests that were comparable to the tests by Luzzi *et al.* (2017*b*).

Partial dissociation of voice-identity recognition of newly-learned and familiar voices

The present findings support a critical involvement of right posterior/mid temporal lobe structures particularly for recognizing newly-learned unfamiliar, in comparison to familiar, voices. This partial dissociation between newly-learned unfamiliar and familiar voice-identity processing is in accordance with previous reports on dissociations of unfamiliar and familiar voice processing in autism spectrum disorders (Schelinski *et al.*, 2016*b*) and in lesion case reports (Van Lancker *et al.*, 1988, 1989). Such potential dissociation may, in our view, be explained by the prototype model of voice-identity processing (Lavner *et al.*, 2001; for review Maguinness *et al.*, submitted for publication). We speculate that highly familiar voices can be readily recognized as the incoming voice is compared to a previously established voice-reference pattern. In contrast, for newly-learned unfamiliar voices such a voice-reference pattern may not yet be established (Fontaine *et al.*, 2017). We suppose that right posterior/mid temporal lobe structures may support the acoustical voice-identity feature analysis that is necessary to establish a new voice-specific reference pattern and/or may be critical for recognition when such a reference pattern is not yet available. The involvement of the right posterior/mid temporal lobe in such acoustical analysis is in agreement with neuroimaging studies suggesting that this region analyses complex spectro-temporal voice features (von Kriegstein and Giraud, 2004; Andics *et al.*,

2010; Roswandowitz *et al.*, 2017). Conversely, familiar voice-identity recognition abilities (i.e. familiarity decision and semantic association) were associated with lesions in the left posterior temporal lobe. This finding, however, should be interpreted with caution as it did not survive correction for lesion volume and hearing level and was also not present when patients with traumatic brain injury were excluded from the analysis. To our knowledge, a particular involvement of the left posterior temporal lobe has not been reported in studies explicitly testing familiar voice-identity recognition (Van Lancker *et al.*, 1988, 1989; von Kriegstein and Giraud, 2004; Birkett *et al.*, 2007; Bethmann *et al.*, 2012) and we also had no explicit *a priori* hypothesis on this region. However, several meta-analysis neuroimaging studies and patient lesion reports suggest a role of the left posterior temporal lobe in name recognition and semantic information retrieval (Blank *et al.*, 2014*a*; Liebenthal *et al.*, 2014; Vitali *et al.*, 2015). It is conceivable that lesions in such a region would lead to deficits in familiar voice-identity recognition tasks.

Conclusions

Our study substantially contributes towards resolving the discrepancy between neuroimaging and lesion reports on voice-identity recognition. The findings have important implications for current voice-identity processing models. They support the central assumption that the right posterior/mid temporal lobe is crucially involved in voice-identity recognition alone and not to the same extent in face-identity recognition or speech processing. However, they also uncover two novel aspects of voice-identity recognition that are not yet included in current models: (i) the critical involvement of the right inferior parietal lobe, particularly the supramarginal gyrus, in voice-identity recognition tasks requiring voice-face association; and (ii) a potential partial dissociation of identity recognition for voices with varying levels of familiarity.

Acknowledgements

Funding

Supplementary material

Supplementary material is available at *Brain* online.

References

Alba-Ferrara L, Weis S, Damjanovic L, Rowett M, Hausmann M. Voice identity recognition failure in patients with schizophrenia. J Nerv Ment Dis 2012; 200: 784–90.

Andics A, Gácsi M, Faragó T, Kis A, Miklósi Á. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr Biol 2014; 24: 574–8.

Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z. Neural mechanisms for voice recognition. Neuroimage 2010; 52: 1528–40.

Ashburner J, Friston KJ. Voxel-based morphometry–the methods. Neuroimage 2000; 11(6 Pt 1): 805–21.

Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005; 26: 839–51.

Assal G, Buttet J, Thuillard F. Cerebra hemispheres and auditory perception. Rev Med Suisse Romande 1981; 101: 177–85.

Assal G, Zander E, Kremin H, Buttet J. Voice discrimination in patients with cerebral cortical lesions [in French]. Schweiz Arch Neurol Neurochir Psychiatr 1976; 119: 307–15.

Baldo JV, Arevalo A, Patterson JP, Dronkers NF. Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. Cortex 2013; 49: 658–67.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, et al. Voxel-based lesion-symptom mapping. Nat Neurosci 2003; 6: 448–50.

Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 2004; 8: 129–35.

Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 2003; 14: 2105–9.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403: 309–12.

Bethmann A, Scheich H, Brechmann A. The temporal lobes differentiate between the voices of famous and unknown people: an event-related fMRI study on speaker recognition. PLoS One 2012; 7: e47626.

Binder JR, Pillay SB, Humphries CJ, Gross WL, Graves WW, Book DS. Surface errors without semantic impairment in acquired dyslexia: a voxel-based lesion-symptom mapping study. Brain 2016; 139: 1517–26.

Birkett PB, Hunter MD, Parks RW, Farrow TF, Lowe H, Wilkinson ID, et al. Voice familiarity engages auditory cortex. Neuroreport 2007; 18: 1375–8.

Blank H, Anwander A, von Kriegstein K. Direct structural connections between voice- and face-recognition areas. J Neurosci 2011; 31: 12906–15.

Blank H, Kiebel SJ, von Kriegstein K. How the human brain exchanges information across sensory modalities to recognize other people. Hum Brain Mapp 2014a; 36: 324–39.

Blank H, Wieland N, von Kriegstein K. Person recognition and the brain: merging evidence from patients and healthy individuals. Neurosci Biobehav Rev 2014b; 47: 717–34.

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J Neurosci 2014; 34: 4548–57.

Boucher J, Lewis V, Collis G. Familiar face and voice matching and recognition in children with autism. J Child Psychol Psychiatry 1998; 39: 171–81.

Brett M, Leff AP, Rorden C, Ashburner J. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 2001; 14: 486–500.

Campanella S, Belin P. Integrating face and voice in person perception. Trends Cogn Sci 2007; 11: 535–43.

Caspers S, Schleicher A, Bacha-Trams M, Palomero-Gallagher N, Amunts K, Zilles K. Organization of the human inferior parietal lobule based on receptor architectonics. Cereb Cortex 2013; 23: 615–28.

Chandrasekaran B, Chan AH, Wong PC. Neural processing of what and who information in speech. J Cogn Neurosci 2011; 23: 2690–700.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006; 31: 968–80.

Duchaine B, Nakayama K. The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 2006; 44: 576–85.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 2005; 25: 1325–35.

Ellis HD, Jones DM, Mosdell N. Intra- and inter-modal repetition priming of familiar faces and voices. Br J Psychol 1997; 88 (Pt 1): 143–56.

Fine DR. A life with prosopagnosia. Cogn Neuropsychol 2012; 29: 354–9.

Fontaine M, Love SA, Latinus M. Familiarity and voice representation: from acoustic-based representation to voice averages. Front Psychol 2017; 8: 1180.

Formisano E, De Martino F, Bonte M, Goebel R. "Who" is saying "what"? brain-based decoding of human voice and speech. Science 2008; 322: 970–3.

Gainotti G, Barbier A, Marra C. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. Brain 2003; 126(Pt 4): 792–803.

Gainotti G, Ferraccioli M, Quaranta D, Marra C. Cross-modal recognition disorders for persons and other unique entities in a patient with right fronto-temporal degeneration. Cortex 2008; 44: 238–48.

Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, et al. Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 2009; 47: 123–31.

Gaudrain E, Li S, Ban V, Patterson R. The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Interspeech 2009; 1: 152–5. doi: 10.6084/m9.figshare.870509.

Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia 2010; 48: 1104–14.

Hailstone JC, Ridgway GR, Bartlett JW, Goll JC, Buckley AH, Crutch SJ, et al. Voice processing in dementia: a neuropsychological and neuroanatomical analysis. Brain 2011; 134: 2535–47.

Härting C, Markowitsch HJ, Neufeld H, Calabrese P, Deisinger K, Kessler J. Wechsler Gedächtnis Test-Revidierte Fassung (WMS-R): Deutsche adaptation der revidierten Fassung der Wechsler memory scale. Göttingen: Hogrefe; 2000.

Holig C, Focker J, Best A, Roder B, Buchel C. Activation in the angular gyrus and in the pSTS is modulated by face primes during voice recognition. Hum Brain Mapp 2017; 38: 2553–65.

Jiahui G, Garrido L, Liu RR, Susilo T, Barton JJ, Duchaine B. Normal voice processing after posterior superior temporal sulcus lesion. Neuropsychologia 2017; 105: 215–22.

Kimberg DY, Coslett HB, Schwartz MF. Power in Voxel-based lesion-symptom mapping. J Cogn Neurosci 2007; 19: 1067–80.

Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG. Voice recognition in aphasic and non-aphasic stroke patients. J Neurol 2009; 256: 1303–6.

Lavner Y, Gath I, Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun 2000; 30: 9–26.

Lavner Y, Rosenhouse J, Gath I. The prototype model in speaker identification by human listeners. IJST 2001; 4: 63–74.

Liebenthal E, Desai RH, Humphries C, Sabri M, Desai A. The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. Front Neurosci 2014; 8: 289.

Luzzi S, Baldinelli S, Ranaldi V, Fabi K, Cafazzo V, Fringuelli F, et al. Famous faces and voices: differential profiles in early right and left semantic dementia and in Alzheimer's disease. Neuropsychologia 2017a; 94: 118–28.

Luzzi S, Coccia M, Polonara G, Reverberi C, Ceravolo G, Silvestrini M, et al. Selective associative phonagnosia after right anterior temporal stroke. Neuropsychologia 2017b. [Epub ahead of print]. pii: S0028-3932(17)30182-3. doi: 10.1016/j.neuropsychologia.2017.05.016.

Macmillan NA, Creelman CD. Detection theory: a user's guide. New York; London: Psychology Press; 2004.

Maguinness C, Roswandowitz C, von Kriegstein K. (in revision) Understanding the mechanisms of familiar voice recognition in the human brain Neuropsychologia (Special Issue: Familiar Voices).

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 2001; 356: 1293–322.

Mou X, Bai F, Xie C, Shi J, Yao Z, Hao G, et al. Voice recognition and altered connectivity in schizophrenic patients with auditory hallucinations. Prog Neuropsychopharmacol Biol Psychiatry 2013; 44: 265–70.

Neuner F, Schweinberger SR. Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn 2000; 44: 342–66.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971; 9: 97–113.

Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PE, et al. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage 2015; 119: 164–74.

Perrodin C, Kayser C, Abel TJ, Logothetis NK, Petkov CI. Who is that? Brain networks and mechanisms for identifying individuals. Trends Cogn Sci 2015; 19: 783–96.

Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice cells in the primate temporal lobe. Curr Biol 2011; 21: 1408–15.

Petkov CI, Logothetis NK, Obleser J. Where are the human speech and voice regions, and do other animals have anything like them? Neuroscientist 2009; 15: 419–29.

Rogalsky C, Poppa T, Chen KH, Anderson SW, Damasio H, Love T, et al. Speech repetition as a window on the neurobiology of auditory-motor integration for speech: a voxel-based lesion symptom mapping study. Neuropsychologia 2015; 71: 18–27.

Rorden C, Brett M. Stereotaxic display of brain lesions. Behav Neurol 2000; 12: 191–200.

Roswandowitz C, Maguinness C, von Kriegstein K. Deficits in voice-identity processing: acquired and developmental phonagnosia. In: Frühholz S, Belin P, editors. The oxford handbook of voice perception. Oxford: Oxford University Press; In press.

Roswandowitz C, Mathias Samuel R, Hintz F, Kreitewolf J, Schelinski S, von Kriegstein K. Two cases of selective developmental voice-recognition impairments. Curr Biol 2014; 24: 2348–53.

Roswandowitz C, Schelinski S, von Kriegstein K. Developmental phonagnosia: linking neural mechanisms with the behavioural phenotype. Neuroimage 2017; 155: 97–112.

Schelinski S, Borowiak K, von Kriegstein K. Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. Soc Cogn Affect Neurosci 2016a; 11: 1812–22.

Schelinski S, Roswandowitz C, von Kriegstein K. Voice identity processing in autism spectrum disorder. Autism Res 2016b; 10: 155–68.

Schmidt KH, Metzler P. Wortschatztest (WST). Göttingen; Hogrefe; 1992.

Schuri U, Benz R. Gesichter-Namen-Lerntest. Frankfurt am Main: Swets & Zeitlinger B.V.; 2000.

Sharp DJ, Scott G, Leech R. Network dysfunction after traumatic brain injury. Nat Rev Neurol 2014; 10: 156–66.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; 23 (Suppl 1): S208–19.

Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. Neuropsychologia 1987; 25: 829–34.

Van Lancker DR, Canter JG. Impairment of voice and face recognition in patients with hemispheric damage. Brain Cogn 1982; 1: 185–95.

Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH. Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex 1988; 24: 195–209.

Van Lancker DR, Kreiman J, Cummings J. Voice perception deficits: neuroanatomical correlates of phonagnosia. J Clin Exp Neuropsychol 1989; 11: 665–74.

Vitali P, Rouleau I, Deschaintre Y, Mina D, Brazeau M, Lanthier S, et al. Proper name anomia in poststroke aphasics: evidence from a multiple-case study. Neurocase 2015; 21: 563–72.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 2003; 17: 48–55.

von Kriegstein K, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 2004; 22: 948–55.

von Kriegstein K, Giraud AL. Implicit multisensory associations influence voice recognition. PLoS Biol 2006; 4.

von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. J Cogn Neurosci 2005; 17: 367–76.

Warren J, Scott S, Price C, Griffiths T. Human brain mechanisms for the early analysis of voices. Neuroimage 2006; 31: 1389–97.

Wechsler D. Wechsler Memory Scale—Third Edition (WMS–III) administration and scoring manual. San Antonio, TX: The psychological corporation; 1997.

Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W, et al. Connected speech production in three variants of primary progressive aphasia. Brain 2010; 133: 2069–88.

Yardley L, McDermott L, Pisarski S, Duchaine B, Nakayama K. Psychosocial consequences of developmental prosopagnosia: a problem of recognition. J Psychosom Res 2008; 65: 445–51.

Zimmermann P, Fimm B. Testbatterie zur Aufmerksamkeitsprüfung: TAP, Version 2.2. Herzogenrath, Germany: Psytest; 2009.