Voice-Identity Processing in Patients with Brain Lesions

Dissertation zur Erlangung des akademischen Titels

Dr. med.

an der Medizinischen Fakultät der Universität Leipzig

eingereicht von Claudia Dominica Kappes, geboren am 29.06.1988 in Münster

angefertigt am Max-Planck-Institut für Kognitions-und Neurowissenschaften Leipzig und der Tagesklinik für kognitive Neurologie des Universitätsklinikums Leipzig

Betreuer: Prof. Dr. med. Hellmuth Obrig und Prof. Dr. med. Katharina von Kriegstein

Beschluss über die Verleihung des Doktorgrades vom: 14.11.2017

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AIG= anterior insula gyrus
AnG = angular gyrus
AOS = anterior occipitial sulcus
BOLD = blood oxygen level dependent
CoI = cortex of insula
CsI = cingulate sulcus of insula
EEG = electroencephalography
FDG = fluordeoxyglucose
FFG = fusiform gyrus
fMRI = functional magnetic resonance imaging
FRU = face recognition units
FTLD = frontotemporal lobar degeneration
GPR = glottal pulse rate
Hipp = hippocampus
ICH = intracerebral hemorrhage
IFG= inferior frontal gyrus
IOG = inferior occipital gyrus
IPS= intraparietal sulcus
ITG= inferior temporal gyrus
JND = just noticeable difference
LGI = lateral gyrus of Insula
LOG = lateral occipital gyrus
MEG = magnetoencephalography
mFG = midfrontal gyrus
MOG = midoccipital gyrus
MTG = middle temporal gyrus
NPS = neuropsychology
NRU = name recognition units
OS= orbital sulcus
pCI= posterior cortex of insula
peCG = precentral gyrus
peCS = precentral gyrus
PET = positrone emission tomography
PIN = person identity nodes
PL = parietal lobe
pMTG = midtemporal gyrus
poCG = postcentral gyrus
poCS = postcentral sulcus
POG = posterior occipital gyrus
PS = parietal sulcus
pSTG = posterior superior temporal gyrus
Put = putamen
RO= rolandic operculum
RS = rolando sulcus
SAH = subarachnoid hemorrhage
SD = semantic dementia
SD = standard deviation
SFG= superior frontal gyrus
SFS= superior frontal sulcus
SMG= superior marginal gyrus
sMRI= structural magnetic resonance imaging
SPL= superior parietal pole
STG= superior temporal gyrus
STP= superior temporal pole
Stri= striatum
STS = superior temporal sulcus
Thal = thalamus
VBM = voxel based morphometry
VRU = voice recognition units
VTL = vocal tract length

1 INTRODUCTION

1.1 Why Research on Voice-Identity Processing in a Medical Context?

The human voice is an important carrier of information as it conveys speech but also paralinguistic features like a speaker's identity and affective information (Belin et al., 2004; Kreiman and Sidtis, 2011). In everyday life, we can tell whether a person we speak to on the phone is familiar to us without effort. Listeners can determine vocal information such as the gender, approximate age, speaker size and even attractiveness, all of which individuate a speaker's voice (Lass et al., 1976; Collins and Missing, 2003; Ives et al., 2005; von Kriegstein et al., 2006; Xu et al., 2013).

Brain injury can lead to various deficits including communication impairment. Deficient speech as a linguistic entity has been well investigated in brain lesioned patients, e.g. the condition of aphasia. Aphasia includes impaired speech production and comprehension (Kertesz and McCabe, 1977; Brownsett et al., 2014). However, voice processing has not been studied as extensively until today, although it is of great relevance for human interaction on the non-linguistic level. The human voice can be considered an auditory equivalent to facial expression. Our brain is capable of extracting not only the meaning of speech, but also personal information about the speaker. For this reason, the term "auditory face" (Belin et al., 2004) has been established. Not only do we know whether a person is familiar to us or not, but we can also tell by their prosody if the speaker is furious, cheerful, nervous, sad or tired, etc. We can also determine if they are native or nonnative speakers and can tell which region of a country they are from. Their voice gives us hints on whether the speaker is young or old, male or female, and can even provide an idea of one's health status (Kreiman and Sidtis, 2011). The human voice already plays a role very early in development. Neonates prefer their mother's voice over others which indicates intrauterine familiarization with voices. This way, our voice has an influence on the early development of bonding mechanisms (DeCasper and Fifer, 1980; Kisilevsky et al., 2003). For an overview of important paralinguistic features delivered through the human voice see table 1.

Some kinds of judgments listeners make from hearing a voice:

Physical characteristics of the speaker
Age
Appearance (height, weight, attractiveness)
Dental/oral/nasal status
Health status, fatigue
Identity
Intoxication
Race, ethnicity
Sex
Sexual orientation
Smoker/non-smoker
Psychological characteristics of the speaker
Arousal
Competence
Emotional status/mood
Intelligence
Personality
Psychiatric status
Stress
Truthfulness
Social characteristics of the speaker
Education
Occupation
Regional origin
Role in conversational setting
Social status

Table 1: Judgments made from voice (Kreiman and Sidtis, 2011), page 2

Previous clinical studies indicate that voice- identity processing can be affected through brain injury (Assal et al., 1976; Van Lancker and Canter, 1982; Neuner and Schweinberger, 2000; Lang et al., 2009; Hailstone et al., 2011). This deficit is called phonagnosia (van Lancker, 1982). Phonagnosia is defined as a deficit in voice-identity processing including the analysis of individuating acoustical voice features, the recognition of familiar voices and the association of semantic information to a recognized voice. Phonagnosia can be present after brain damage, i.e. acquired phonagnosia or as an inborn condition, i.e. developmental phonagnosia (Garrido et al., 2009; Roswandowitz et al., 2014; Xu et al., 2015). Losing the ability to recognize the identity of people through their voice is irritating to those affected, but it is also

disruptive to their own environment, as it is our ability to read voices which forms an important component of human interaction.

People with phonagnosia lack the ability to identify a speaker by their voice. Answering a phone call without knowing who is talking on the other end would be an everyday life experience for patients affected. Hearing voices in other rooms and not being able to identify them is another situation that is likely to occur. The prevalence of developmental phonagnosia ranges from 0.1-3% (Roswandowitz et al., 2014; Shilowich and Biederman, 2016) while in a study by Neuner and Schweinberger 13 out of 36 patients after brain damage showed a voice recognition deficit (Neuner and Schweinberger, 2000). Therefore, it seems likely that phonagnosia is not a rare condition and is worth taking a systematical look at.

In the following chapters I will provide an overview on the foundations of research on voice recognition.

1.2 Theory of Voice Production

To investigate voice recognition mechanisms, a crucial question to answer is: Which voice features make a human voice unique?

According to the source filter theory (Fant, 1960), the human vocal tract comprises a sound source and a filter that contribute to the formation of individual voice quality. The source of the vocal sound is located in the larynx, the glottal folds. The vibration and movement of the glottal folds result in vocalization. When the glottal folds, at the end of the vibration, close completely, a voiced sound is produced whereas glottal folds in a paramedian position produce whispered speech. The glottal pulse rate (GPR), meaning the rate of vibration of the glottal folds during articulation, determines a speaker's fundamental frequency (f0). High glottal pulse rates lead to higher, low glottal pulse rates to lower frequencies. The fundamental frequency is perceived by the listener as a speaker's pitch.

The vocal tract includes the supralaryngeal area ascending to mouth and tip of the nose and operates as a filter of the sound wave created by the glottal folds. Individual size and shape of a vocal tract determine the speaker's formant frequencies. Formants are

defined as the spectral peaks of the sound spectrum of the human voice (Fant, 1960). Each spoken vowel and consonant has its own fundamental frequency spectrum; the formant with the lowest frequency is called the first formant (f1).

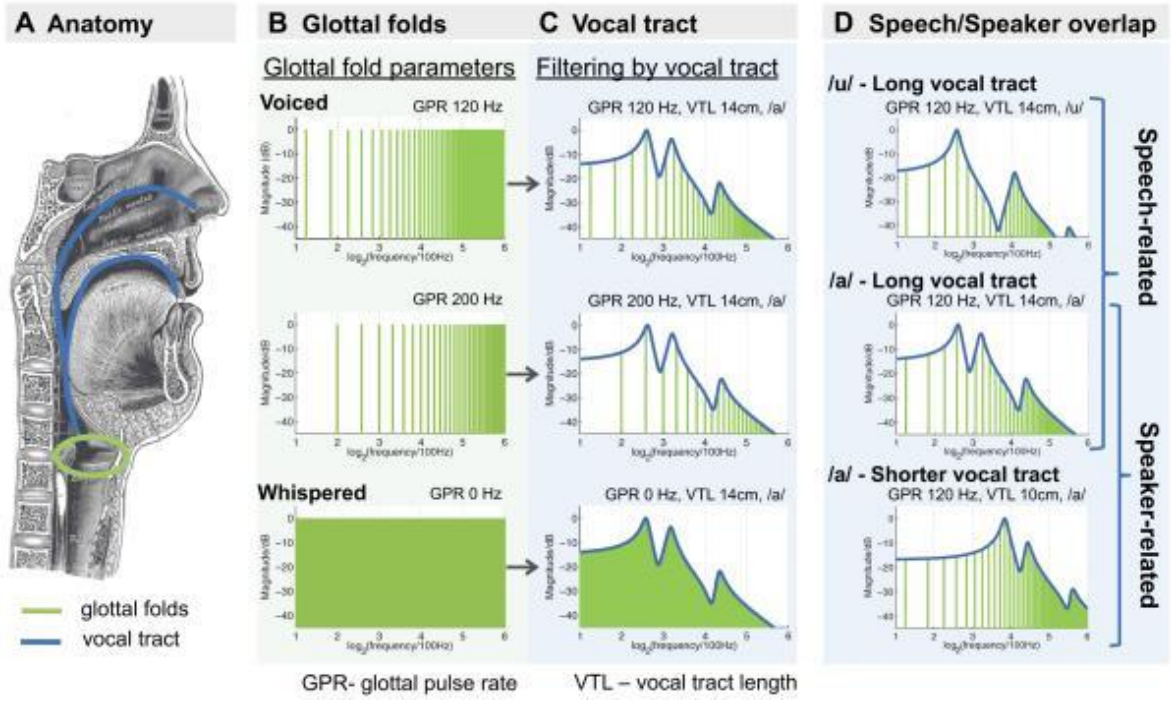


Figure 1: The contribution of glottal fold and vocal tract parameters to the speech output (von Kriegstein et al., 2010). **A)** Anatomy of the vocal tract and location of vocal fold, sagittal section **B)** Different sounds determined by glottal fold parameters: Vibration of glottal folds results in lower voice (GPR 120 Hz) or higher voice (GPR 200 Hz), constricted vocal folds produce whispered speech (GPR 0 Hz). **C)** Blue lines represent formants (amplitude peaks at certain frequencies through the filtering in the vocal tract). Different glottal fold parameters do not influence formant position. **D)** Speech- and speaker-related vocal tract parameters influence formant position. Formant shifts are demonstrated with speech sounds /u/ and /a/ and with a shorter and longer vocal tract. Figure taken with kind permission from the Journal of Neuroscience (von Kriegstein et al., 2010).

Different vocal tract lengths (VTL), along with the mass and the length of the vocal folds, add to individual voice quality. Males tend to have longer vocal folds than females that vibrate at lower frequencies and therefore result in a lower voice sound. Stiffness of the vocal folds causes higher vibration rates and, as in a string instrument with the strings tuned higher, it leads to a higher pitch. The average fundamental frequency (f0) is about 115 Hz for men and 220 Hz for women (Kreiman and Sidtis, 2011).

Pitch (i.e. GPR) and Timbre (i.e. VTL) were identified to be relevant auditory cues to individuate a human voice in various studies (for review see Mathias and von Kriegstein, 2014). GPR has been proven to be an important vocal cue for the discrimination and similarity judgment of unfamiliar speakers (Singh and Murry, 1978; Baumann and Belin, 2010). In a study by Gaudrain and colleagues, VTL was found to be a consistent cue of voice quality when making discriminative judgments on a speaker whose GPR and VTL had been varied artificially (Gaudrain et al., 2009). Pitch (GPR) also appears to play an essential role in sex and gender discrimination (Skuk and Schweinberger, 2013). Further studies discovered that both GPR as well as VTL are required for recognizing the sex of a speaker (Smith and Patterson, 2005).

One approach to investigate the influence of acoustical voice features to voice-identity processing is to systematically modify those features. For instance, modern software (STRAIGHT (A Kawahara et al., 1999)) allows for manipulation of GPR and VTL for experimental purposes. It facilitates the investigation of their influence on speaker recognition, for example, one voice's GPR and VTL can be varied to an extent that the listener can no longer tell voices apart.

Next to these two parameters, there are more aspects of voice production that have an impact on the voice quality perceived by the listener, like for example lip, tongue and jaw position. Individual speaking habits, intonation or accent, influence a speaker's fundamental frequency (Mathias and von Kriegstein, 2014). A voice can sound breathy or whispery or hoarse. These voice qualities are caused by different glottal settings, a gradual closure of the vocal folds. However, these latter qualities remain difficult to assess objectively (Laver, 1980). Intriguingly, even though the human voice may vary in so many aspects, humans are still remarkably precise in correctly recognizing who is speaking.

1.3 Models of Person Recognition

How are voices processed on a cognitive and neuronal basis? A theoretical model of face recognition that was proposed in 1986 by Bruce and Young (Bruce and Young, 1986a) has strongly influenced current models of voice recognition. This face recognition model was first extended to voice recognition by Ellis et al. in 1997 (Ellis et al., 1997). According

to this model, voice and face information are processed in segregated pathways up to the level of person identity nodes (PINs). PINs provide access to multi-modal identity-specific semantic codes of a person familiar to us. When person-specific knowledge is extracted from PINs, e.g. phonological codes can be activated to retrieve and produce someone's name. A familiarity feeling for a known person is created on a modality specific level (face - and voice recognition units).

In 2004, Belin and colleagues proposed a modified model for voice recognition based on the described face recognition model (Belin et al., 2004). Speech, vocal affect and voice identification are functionally dissociated pathways and follow basic structural encoding of auditory cues (see Fig. 2). These three pathways interact with the corresponding pathways of the face processing system but are generally segregated. Voice recognition units (VRUs) form the counterpart to face recognition units (FRUs). The modality specific information converges in Person Identity Nodes.

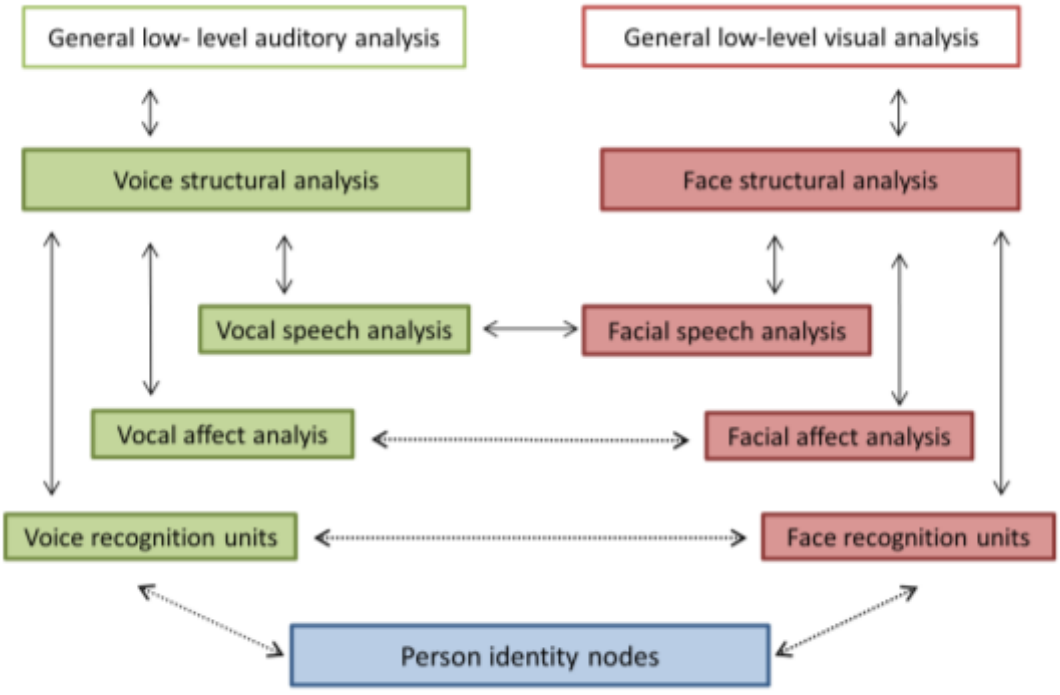


Figure 2: Adapted model on voice and face recognition based on the original model by Bruce and Young: Voice processing mechanisms were adapted from the original model by Belin and colleagues (Belin et al., 2004). Continuous arrows indicate modality-specific interactions, dashed arrows indicate bidirectional interactions.

Dysfunction at different stages results in distinct deficit patterns in voice-identity processing. Recent models imply that impairment at an early, perceptual analysis stage leads to apperceptive deficits. Dysfunction at a higher analysis level (i.e. VRU and PIN) causes associative deficits (i.e. in familiarity decision and semantic association) (Roswandowitz C, Maguinness C, v Kriegstein K., under rev.). Clinical and neuroimaging evidence from the past years has questioned the existence of modality-free PINs with different competences for each hemisphere (Gainotti, 2007). Other studies found evidence for early interaction of visual and auditory information areas during processing (Blank et al., 2011; Schall et al., 2013).

However, neuronal representations of voice-identity processing have not been fully understood until today. In the following, I introduce the current state of the art in voice-identity processing by reviewing neuroimaging and clinical studies.

1.4 Neuronal Mechanisms of Voice-Identity Processing

Belin and colleagues first characterized voice-selective areas in the human cortex (Belin et al., 2000) employing fMRI. Healthy participants listened to vocal versus non-vocal sounds in the MRI. In their study, the upper bank of the central part of the superior temporal sulcus (STS) showed bilateral voice-selective activation, although activation was stronger in the right hemisphere. The involvement of the right STG/S region in voice-identity recognition remained a consistent finding in subsequent studies using different designs (Shah et al., 2001; Belin and Zatorre, 2003; Kriegstein and Giraud, 2004; Formisano et al., 2008; Mathias and von Kriegstein, 2014). Figure 3 shows repeatedly found voice-selective brain response along the right STG/S.

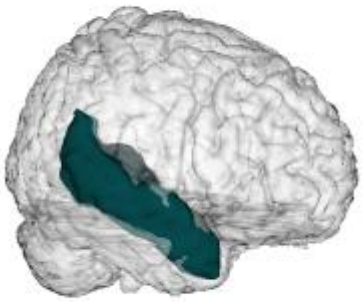


Figure 3: Voice- selective activation along the right superior temporal gyrus and sulcus as a common finding in fMRI studies (Belin et al., 2004; Kriegstein and Giraud, 2004).

Belin and colleagues reported voice-sensitive brain response when participants passively listened to vocal as compared to non-vocal sounds, while other studies used different task designs. This localizer allows for an investigation of acoustical voice processing and evokes high brain responses. One drawback of this localizer is the passive listening design making it hard to control for an attentional bias during the experiment. Also, it is not specifically assessing voice-identity processing as the presented sounds included emotional as well as speech information. A later study by von Kriegstein et al. developed a new approach to locate brain areas sensitive to voice identity. They directed participants' attention in the MRI either to vocal identity or speech information of a target sentence in an active listening task (von Kriegstein et al., 2003). Participants had to choose either the target voice or the verbal content while listening to the same auditory stimulus. In turn, this test design allows the investigation of brain areas which specifically respond to either voice-identity or speech information dependent on the task and irrespective of the auditory stimulus. Increased BOLD response was found in the right anterior STS when listeners focused on the voice-identity information in a target phrase, while verbal information elicited activation in the left posterior and middle temporal region. A follow-up study with the same task showed that activation in the posterior part of the right STS was stronger when unfamiliar voices had to be recognized while the anterior STS was responsive during both unfamiliar and familiar voice recognition (Kriegstein and Giraud, 2004). In the same vein, the authors reported

different functional connectivity patterns dependent on the voice familiarity. During unfamiliar voice recognition, the right posterior STS interacted with the mid- and anterior STS in the right hemisphere as well as the posterior and mid/anterior STS regions in the left hemisphere. During familiar voice recognition, the right anterior STS was functionally connected with prefrontal and amygdala/para-/hippocampal regions in the right hemisphere.

Next to fMRI studies, voice-sensitive processing in the human brain was also evidenced by EEG and MEG investigations. These studies revealed voice-specific electrophysiological responses at a peak of around 200 ms in comparison to other sounds (Levy et al., 2003; Charest et al., 2009; Capilla et al., 2013; Schall et al., 2015). Furthermore, there is one neurostimulation study applying tDCS. Stimulation of the right posterior temporal lobe induced deficits in vocal sound detection in healthy participants (Bestelmeyer et al., 2011).

Most current models on voice recognition are based on findings from studies in healthy participants. However, clinical evidence is crucial to test predictions made by investigations on healthy participants, because clinical evidence allows for causal interpretations on neuronal processes and behavior.

1.5 Clinical Studies

This chapter will provide an overview of clinical studies investigating voice processing (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Van Lancker et al., 1988, 1989; Peretz et al., 1994; Neuner and Schweinberger, 2000; Lang et al., 2009; Hailstone et al., 2010, 2011; Luzzi et al., 2017a). The aforementioned studies explored voice-identity processing by means of behavioral testing and were aimed at unraveling the neuroanatomical correlates of voice-identity processing deficits in brain injured patients.

Van Lancker and Canter published the first study on famous face and voice recognition in 1982. They tested 21 left- and 9 right-brain damaged patients by showing them photographs and sound recordings of famous personalities (politicians and entertainers). All left-brain damaged patients were aphasic. Both face and voice recognition impairments were more frequent in the right-hemisphere group and showed a tendency

to co-occur. The most relevant finding was that there were also dissociations between face and voice recognition impairment as one patient with a right-hemipsheric lesion showed a selective impairment in familiar voice recognition (Van Lancker and Canter, 1982).

In 1989, van Lancker and colleagues (Van Lancker et al., 1989) further investigated voice-identity processing in brain damaged patients by focusing on the dissociation between unfamiliar voice discrimination and familiar voice recognition (apperceptive vs. associative processing). 56 patients were included (25 left- and 24 right-hemisphere damaged, 5 bilateral). Patients were matched with 48 healthy controls. 43 patients obtained CT scans that were used to classify lesion sites within the hemispheres. The results showed a correlation of right parietal lobe damage with a poor performance in famous voice recognition, whereas unfamiliar speaker discrimination was impaired in right- and left-temporal lobe damaged patients. Also, they found four patients with temporal lobe lesion and preserved task performance. These patients had lesions exclusively in the left hemisphere indicating a higher relevance of the right hemisphere during unfamiliar voice discrimination.

Van Lancker et al. used CT scans in their second study to classify lesion sites either to frontal, parietal, temporal and occipital and, if more lobes were affected, they used combined classifications (fronto-parietal etc.). Imaging technologies at that time did not allow for more detailed anatomical analysis.

Studies by van Lancker et al. in brain damaged patients provide evidence that the right hemisphere is essential for speaker recognition. Moreover, they implied that associative voice-processing impairment occurs with right parietal lobe damage and apperceptive deficits are associated with temporal lobe damage in both hemispheres.

A decade later, Neuner and Schweinberger tested 36 patients with brain lesions and 20 healthy control subjects with a test battery that comprised tests for face, voice and name recognition. The aim was to detect the dissociation of impairments in the different modalities with an adaptive test design. Patients who showed difficulties in one of the modalities completed further experiments to investigate whether the deficit was modality-specific. These experiments consisted of corresponding objects to each

modality, e.g. voices corresponded with environmental sounds, faces corresponded with pictures of objects and names corresponded with written words (Neuner and Schweinberger, 2000). 1 out of 16 left-brain damaged and 4 out of 13 right-brain damaged patients presented selective voice recognition disorders. One of the right-hemispheric patients showed additional famous face recognition impairment.

This study was the first to describe selective voice recognition deficits (phonagnosia) in a large patient sample. By employing their corresponding auditory object recognition tests they excluded the possibility that voice recognition deficits were caused by a more global auditory deficit. Moreover, they confirmed van Lancker's finding that a voice recognition disorder can occur without an additional voice discrimination deficit.

In 2009, Lang et al. published a patient study on speaker-identity recognition in aphasic and non-aphasic stroke patients, hypothesizing that left-hemisphere damage might lead to aphasia but not necessarily to deficits in speaker identification (Lang et al., 2009). They postulated that right-hemisphere damage without aphasia, on the other hand, might result in selective voice recognition deficits. One task on familiar voice recognition was conducted with 20 patients (11 left- and 9 right-brain damaged) and 17 healthy controls. All patients had brain lesion due to ischemic stroke. Right-brain damaged patients obtained significantly lower scores on the voice recognition task than both left-hemisphere patients and healthy controls. Left-hemisphere damaged patients and healthy controls did not differ significantly in their speaker identification performance. The study provides implications for a dissociation of voice-identity and speech processing deficits. However, language abilities were not tested in right-brain damaged patients, therefore a double-dissociation cannot be excluded.

Another extensive study on voice recognition in patients was provided by Hailstone et al in 2011. This study employed voxel based morphometry (VBM) analysis of MRI scans. Patients with either Alzheimer's disease (n= 22) or frontotemporal lobar degeneration (FTLD) (n=14) were tested on voice perception and recognition, and compared with 35 healthy controls. In order to control for the specificity of deficits, tests of face and name recognition were included in the test battery. The authors hypothesized distinct profiles of voice recognition deficits in both dementia forms, with FTLD presenting more associative (i.e., semantic association to a voice) and Alzheimer's more apperceptive

(i.e., acoustical voice analysis) impairment. They furthermore assumed that voice processing deficits would correspond to anterior temporal lobe atrophy. Perceptual impairments were present in both disease groups but FTLD patients exhibited more severe deficits in voice recognition than the Alzheimer's group. Both disease groups revealed impairment in all other modalities tested, i.e. face and name recognition. This indicated a relative general person recognition deficit rather than a selective deficit in voice-identity processing. VBM analysis revealed that voice recognition as well as face- and name-recognition deficits (cross-modal recognition) were associated with grey matter volume loss in the right temporal pole. A specific region for voice deficits could not be shown. Also, as the study investigated a sample of patients with neurodegenerative disease (Alzheimer's and FTLD), it was difficult to control for the impact of general cognitive deficits on the given task. Both patient groups had performed significantly worse on the neuropsychological assessment than the controls.

The most recent patient study exploring voice- and face recognition deficits was published in 2017 (Luzzi et al., 2017a). The authors employed neuropsychological tests on face and voice naming and recognition in a sample of patients with either Alzheimer's disease (n= 25) or semantic dementia (n= 13) and 34 healthy controls. The aim of the study was to examine differential impairment patterns for person recognition in both disease groups. While AD patients showed impaired performance on the naming tests, SD patients were significantly impaired in both naming and recognition of face and voice. PET scans of 12 SD patients showed strong correlations of face- and voice recognition and a reduced FDG (fluordeoxyglucose) uptake in the right temporal lobe.

Luzzi et al. provided detailed neuropsychological assessment on both face and voice recognition. There was a strong correlation between reduced FDG uptake in PET and face- and voice recognition impairment in the right but not the left temporal lobe. PET was performed on a relatively small sample size (n= 12). Although correlations were stronger for voice recognition impairment than face recognition impairment, a voice-specific region within the right temporal lobe could not be detected with this method.

For an overview of results from the described clinical studies please see table 2. In summary, clinical studies on voice-identity processing provide strong evidence that the right hemisphere plays an essential role for voice-processing. Moreover, there are

indications for an involvement of the temporal lobe. Selective voice-identity processing impairment could not yet be associated with distinct brain areas as most studies described multi-modal person recognition deficits.

Further studies on voice recognition deficits were conducted in the psychiatric area of research, i.e. schizophrenia and Autism Spectrum Disorders. Schizophrenia research aimed to link possible voice identity recognition impairments to the formation of auditory hallucinations, and detected difficulties in recognition. Also, it appeared that tested patients relied on different vocal cues when identifying a voice (Alba-Ferrara et al., 2012; Chhabra et al., 2012). Autism Spectrum Disorders are characterized by deficits in communication and social interaction. Recent studies evidenced behavioral deficitis in voice-identity processing in ASD (Schelinski et al., 2016, n.d.). On the neuronal level, this behavioral deficit is reflected by reduced BOLD response in voice-sensitive brain regions along the STG/S (Gervais et al., 2004; Schelinski et al., 2016).

Authors	Patients	Methods	Results
Van Lancker and Canter (1982)	21 Left- 9 Right - hemisphere damaged patients	Famous voice and face recognition	Face and Voice recognition more impaired in right-hemisphere group
Van Lancker et al. (1989)	23 Left- 15 Right- hemisphere damaged patients 48 healthy controls	Famous voice recognition & unfamiliar voice discrimination (UVD)	Famous voice recognition impaired in right-hemisphere patients, UVD in both sides
Neuner and Schweinberger (2000)	36 brain damaged patients 20 healthy controls	Famous names, voice and face recognition and unfamiliar face and voice discrimination	1/16 L and 4/13 R showed specific Famous voice recognition impairment
Lang et al. (2009)	11 Left & 9 Right-hemisphere damaged patients 17 healthy controls	Familiar voice recognition	Famous voice recognition more impaired in right-brain damaged patients
Hailstone et al. (2011)	22 AD patients 14 FTLD patients, 35 healthy controls	Famous voice, face and name recognition Unfamiliar voice discrimination	Voice recognition impairments more severe in FTLD, association with right anterior temporal pole, no specific region for voice deficit
Luzzi et al. (2017)	13 SD, 25 AD patients, 34 healthy controls	Famous face and voice naming and recognition	Naming deficits in AD & SD, face & voice recognition impaired in SD, correlation of face and voice recognition deficit and RTL atrophy, no specific region for voice deficit

Table 2: Overview of patient studies on voice recognition. FVR: Famous voice recognition, UVD: Unfamiliar voice discrimination, AD= Alzheimer's disease, FTLD: Frontotemporal lobar degeneration, SD= Semantic dementia, RTL = Right temporal lobe

2 RESEARCH QUESTION

In consideration of the mentioned methodological shortcomings of the past studies, our aim was to design a patient study that would address these limitations. They mainly consisted in weak proof for modality-specific deficits and their neuronal representations.

The aim of the present study was to assess selective voice-identity processing deficit in brain damaged-patients and its neuroanatomical representation. We developed a comprehensive behavioral test battery including tests on voice-identity processing as well as face recognition. The voice processing tests included tests on newly-learned voice recognition and familiar voice recognition. We also tested patients in their ability to process acoustical voice features relevant for identity recognition, i.e. vocal pitch and vocal timbre discrimination. Patients with unilateral brain lesions were chosen as a target disease group to ensure clearly defined lesion locations in structural MRI scans. We additionally conducted extensive neuropsychological assessment to eliminate a possible influence of cognitive impairment on behavioral test performance. The test battery was designed based on current cognitive models of voice recognition (Belin et al., 2004; von Kriegstein et al., 2006). To assess neuroanatomical correlates of voice processing deficits, Voxel-based Lesion Symptom Mapping (Bates et al., 2003a) was applied in a further analysis which is not part of the current dissertation work.

By combining a large patient sample, a comprehensive behavioral test battery, and lesion analysis through VLSM we aimed for insight into neuronal mechanisms of human voice processing. For the behavioral data analysis, we hypothesized that voice recognition deficits were more likely to occur in patients with right-hemispheric lesions than with left-hemispheric lesions.

According to the hypotheses we deducted the following research questions:

1. Do voice recognition deficits occur more frequently in patients with lesion in the right hemisphere as compared to patients with lesion in the left hemisphere?
2. Are voice-identity processing deficits modality-specific, or do they co-occur with face recognition deficits?

3 MATERIALS AND METHODS

3.1 Subject Details

Forty unilaterally left- and right- brain damaged patients participated in the study. All patients were recruited from the "Clinic for Cognitive Neurology" at the University Hospital of Leipzig. 11 patients were enrolled in the Clinic's therapy program at the time of testing and 29 patients were selected from the Clinic's database and invited via a written letter and a telephone conversation. Written consent was collected from all participants before the experiment. The study design was approved by the Ethics Committee of the University of Leipzig.

Exclusion criteria were severe aphasia, severe cognitive impairment or subjective disability to follow test instructions. No patient was diagnosed with psychiatric disorders as dementia, schizophrenia, personality disorder or depression.

Brain damage was caused by cerebral vascular incident (i.e. ischemia, intracerebral hemorrhage or subarachnoid hemorrhage) or tumor surgery. We did not include traumatic brain injury to avoid inaccuracy in lesion locations through diffuse axonal damage. We recruited 26 patients with ischemic, 5 with intracerebral hemorrhage and 6 patients with subarachnoid hemorrhage. 3 tumor patients with benign low grade astrocytoma (1 oligoastrocytoma and 2 meningioma) after tumor excision were included into the study. All lesions were located unilaterally in either left (n=16) or right hemisphere (n=24) and varied in their location within the hemisphere. Time since onset ranged between 2 and 57 months at the time of testing. 33 patients had received structural brain MRI before their admission to the Clinic.

All patients had hearing abilities allowing for normal speech and sound perception which we assessed via pure-tone audiometry (250 – 8000Hz) using a screening audiometry (MADSEN Micromate 304, GN Otometrics, Copenhagen, Denmark).

3 patients were excluded from the analysis. One patient was excluded because of severe aphasia and one because of severe cognitive impairment. A third patient was excluded because her time since onset was only one month at the time of testing and she reported great difficulties understanding the test instructions. German was mother

tongue of all participants, except for one patient (32 years old) who had grown up in Russia, but had lived in Germany for 11 years at the time of testing and was fluent in German.

14 patients were diagnosed with mild cognitive impairment after surgery or CVA. Mild aphasia was present in 4 patients. Brain damage was unilateral in all patients, but showed variation in location within the hemisphere. In 5 patients brain damage was subcortical only (R= 3, L= 2).

All patients were right-handed. Handedness was assessed with the Edinburgh inventory (Oldfield, 1971).

For detailed demographical data of the patient sample see table 3.

	Patient Sample	**Right-Brain Damage**	**Left-Brain Damage**
Sex	21 F/ 16 M	13 F/ 10 M	8 F/ 6 M
Age Mean *(± SD)* (years)	47.51 (*11.16*)	46.96 (*11.40)*	48.43 (*11.12)*
Age Range (years)	22-63	22-63	29-62
Education *(± SD)* (years)	10.57 (*1.01*)	10.43 (*1.04*)	10.71 (*.99*)
Time post-onset *(±SD)* (months)	24.44 (*15.42*)	27.14 (*14.36*)	20.21 (*16.05*)
Time post- onset range (months)	2-57	7-54	2-57

Table 3: Demographical data of the patient sample: Mean values of age, years of education and time post-onset are depicted. SD = +/ - 1 Standard deviation.

3.2 General Neuropsychological Assessment

To control for cognitive deficits, we conducted a series of neuropsychological tests with each patient.

3.2.1 TAP – "Test of Attentional Performance"

TAP is a test battery that was developed to detect deficits in attentional functions (Fimm and A Zimmermann, 2001). Brain injured patients experience a variable range of deficits in their attentional performance. As attention is no longer considered a singular function, the TAP test battery comprises several subtests that cover different aspects of attentional performance (alertness, covert attention shift, visual scanning etc.). All stimuli are non-verbal and demand reaction through a simple key press. Reaction time and number of mistakes are considered.

We conducted the "alertness" subtest with our patient sample. It assesses reaction time in two conditions: Intrinsic alertness (reaction time to a visual stimulus) and phasic alertness (critical stimulus preceded by a warning signal).

Means for the group were calculated from the normative sample range that considered performance relative to a participant's age. The normative sample contains an age range from 6 to 89 years (N = 1131) with an unknown survey period.

3.2.2 Digit and Spatial Span

Digit and spatial span (Härting et al., 2000) assess auditory and spatial working memory capacity. For the digit span condition, the participant is asked to recall increasing sequences of numbers that are being read out loud by the instructor. The spatial span consists of a board with 9 numbered blocks. Block numbers are hidden to the participant. The instructor tips an increasing sequence of numbers by touching a series of blocks and asks the patient to repeat the sequence by tipping each block he tipped before. We asked patients to repeat the digits and spatial cues in forward order in the first condition and in backward order in the second condition. The longest sequence a participant can repeat correctly is set as his digit and spatial span.

The normative sample contains the age range from 15 to 74 years (N = 210), referring to a survey period from 1996 to 1997. The normative sample is grouped according to age

(seven age groups ranging from 15 to 74 years), gender and school education ('Hauptschule', 'Realschule' and 'Gymnasium').

3.2.3 "Wortschatztest"- Test of German Vocabulary

As a measure for verbal intelligence and language comprehension skills we applied the "Wortschatztest" (Schmidt and A Metzler, 1992). The test contains 42 German words. Each word appears in one line among 5 distractors (fictional German vocabulary) and the participant is asked to mark the target word (i.e. the "real" word) in each line. The words appear in an increasing order of difficulty. The normative sample contains the age range from 16 to 90 years (N = 573). The survey period is unknown.

3.2.4 "Gesichter-Namen-Lerntest"- Face-Name Learning Test

Associative learning capabilities were assessed with the "Gesichter- Namen- Lerntest" (Schuri and Benz, 2000). This test was designed for patients after brain injury that notice deficits in remembering names and faces after the incident. 8 black and white portrait photographs and corresponding names are presented to the patients for 10 seconds who are asked to repeat each name after presentation. All 8 names are disyllabic names. In the second block out of four, photographs are presented in different order and without the names. The patients' task is to name as many photographs as possible. Blocks are repeated until the patient can name all 8 photographs correctly, or until the end of the 4th block. After 30 minutes the patient is asked to recall the names learned before. As a last step, all photographs are presented without names, and the task is again to name as many photographs as possible. Repetition after 30 minutes allows to test not only short term but also intermediate term memory skills. The normative sample contains the age range from 16 to 82 years (N = 76) with an unknown survey period. The normative sample is grouped into six age groups ranging from 15 - 29 to 70 - 85, gender and level of school education ('Hauptschule', 'Realschule' and 'Abitur').

3.3 Peripheral Hearing Assessment

We assessed hearing threshold levels between 250 and 8000 Hz (American National Standards Institute, 2004) with a pure tone auditory screening instrument (MADSEN Micromate 304, GN Otometrics, Copenhagen, Denmark). Testing took place in a soundproof booth in the Clinic.

3.4 Questionnaire

Each patient filled in a questionnaire to estimate the subjective person recognition abilities before and after brain injury. The questionnaire inquired person recognition in general, as well as recognition features (e.g. face, voice, clothing, haircut etc.) the participant might use for person recognition. Patients answered the questions on a scale from 1 to 5, with 1 as very good recognition abilities and 5 as severely affected recognition abilities.

Our aim was to discover the relationship between a patient's self-report and actual test performance (see supplementary material for the questionnaire).

3.5 Behavioral Test Battery

To investigate voice and face recognition mechanisms we employed a comprehensive behavioral test battery. All tests described in the following have been applied in a previous voice processing study conducted by Roswandowitz et al. in 2014. The following test descriptions (chapter 3.5.1- 3.5.4) are quoted from "*Two Cases of selective developmental Voice-Recognition impairments*", Roswandowitz et al., Current Biology, 2014 (Roswandowitz et al., 2014) .

3.5.1 General Procedure

All experiments were carried out on a desktop computer in a soundproof booth. Patients were comfortably seated facing a 21-inch monitor featuring the visual stimuli. Auditory stimuli were presented via Headphones (Sennheiser HD 280 pro, Wennebostel, Germany). The sound level was adjusted individually for each participant. Patients' responses were recorded via keyboard. To ensure that all participants understood the tasks, the experimenter gave oral instructions, in addition to the written instructions prior to each test.

3.5.2 Assessing Voice-Identity Processing Skills

3.5.2.1 The Voice-Face and Voice-Name test: Voice Recognition using Unfamiliar Voices

We assessed voice processing skills in two types of tasks; voice learning of unfamiliar voices and voice recognition of famous voices. In the following section both tasks are

described in detail. The voice-name and voice-face test were employed to investigate learning abilities of unfamiliar voices in our subjects.

3.5.2.2 Stimuli and Presentation Software

The auditory stimulus material was recorded from 14 German native speakers without regional dialect (eight female, age value and SD). All speakers were orally instructed to read sentences in a neutral manner and at a normal speech rate. Each of the six speakers read the same 26 five word declarative sentences (e.g., German: “Der Junge trägt einen Koffer.”, English: The boy carries a suitcase.), and 3 two word declarative sentences (e.g.,”Er sagt.”, He says.). In addition the speakers read 3 five word interrogative sentences (”Trägt der Junge einen Koffer?”, Does the boy carry a suitcase?).

Each speaker read 41 declarative five-word sentences and 5 declarative two-word sentences, as well as 5 interrogative five-word sentences. This resulted in a total set of 714 sentences. High quality auditory recordings were taken in a soundproof recording chamber with a condenser microphone (Rode NT 55 MP; USB Sound Interface: Fast Track MK2, M-Audio, US; 44,1 kHz sampling rate, 16 bit resolution) and the Audacity software (version 1.3.5. beta (http://audacity.sourceforge.net)). The stimuli were post-processed using Audacity (version 1.3.5. beta (http://audacity.sourceforge.net) and Matlab (version 8.1, The MathWorks, Inc., MA, USA), and were normalized for peak amplitude using PRAAT (Boersma and Weenink, 2001).

The visual stimuli comprised pictures of the speakers’ faces. Pictures were recorded with a digital video camera (Legria HF S10 HD-Camcorder, Canon Inc., Japan) and all taken under the same lighting conditions in front of a black background. The speakers’ faces on the pictures were visible from the chin to the hairline. All speakers had a neutral facial expression. No face contained salient visual features such as beards, piercings or glasses. The experiment was implemented in the Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

3.5.2.3 Procedure

Participants learned to associate six unfamiliar voices (three female, three male) with the speaker's face and the speaker's name respectively. Different speakers were used in the two tasks. The experiment lasted around 40 minutes.

Both tests contained a 'female voices' and a 'male voices' block which were identical in structure. The running order of blocks was randomized. Both were divided into four phases of learning and into five testing phases. Within a block, none of the sentences was repeated. Voice-face and voice-name pairs were learned by presenting pairs simultaneously: patients listened to the voices while seeing face or name on the screen. During the first and second learning phase, two cycles each containing five subsequently presented five-word declarative sentences per speaker were presented; the running order of the speakers was randomized. In the third and fourth learning phase, only three sentences per speaker per cycle were introduced. The learning phases were separated by testing phases, assessing the voice-face and voice-name learning performance. The participants listened to an audio sample of one of the speakers and carried out a three-alternative forced choice task, that is, they had to select either the correct face or the correct name associated with the voice at hand. In the first and the second testing phase they received feedback on their decision and the correct voice-face/voice-name pair was presented again. Each of the speakers contributed five sentences to the testing phase. During the third and the fourth testing phase, questions and two-word sentences were employed, respectively, to avoid prosody-driven identity processing. We calculated the learning performance as a mean percent correct over all five testing phases, for each block separately and averaged over both blocks. Participants received feedback in the first second blocks about correct and incorrect response and received correct voice-face pairing. It took approximately 40 minutes to complete one experiment.

3.5.2.4 Voice Recognition using Famous Voices

The test served to assess to what extent patients can recognize voices known from the media. We considered two mechanisms of familiar voice recognition, familiarity decision making and familiar voice naming (semantic association).

3.5.2.5 Stimuli and Presentation Software

The auditory stimulus set contained voice samples of famous (n=40) and unfamiliar (n=20) German speakers. We extracted the voice samples from open access mp3 or wav files available on public radio and television websites. Each sample was five seconds long. The famous voice samples comprised voices from 21 media personalities, eight politicians, seven actors and four musicians (see table 4). In a pilot study, a group of 10 individuals without voice recognition deficits rated the familiarity of a larger stimulus set (56 celebrities) on a scale from 0-5 (very unfamiliar-very familiar). Celebrities rated with an average of three or higher (n=40) were included into the famous voices stimulus set.

The files were edited using the Audacity software (version 1.3.5. beta (http://audacity.sourceforge.net)) and peak amplitude was scaled using the PRAAT software (Boersma and Weenink, 2001). The semantic contents of the auditory samples provided no information about the famous persons' identity or profession. The experiment was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Marcel Reich-Ranicki	Otto Walkes	Till Schweiger
Joachim Löw	Helge Schneider	Anne Will
Günther Jauch	Michael Mittermeier	Alice Schwarzer
Ulrich Wickert	Franz Beckenbauer	Verona Pooth
Hellmuth Karasek	Boris Becker	Heidi Klum
Harald Schmidt	Michael Schumacher	Anke Engelke
Karl Lagerfeld	Gerhard Schröder	Barbara Schöneberger
Alfred Biolek	Helmuth Kohl	Sarah Kuttner
Thomas Gottschalk	Guido Westerwelle	Angela Merkel
Jürgen von der Lippe	Wolfgang Schäuble	Ursula von der Leyen
Oliver Pocher	Erich Honecker	Claudia Roth
Vicco von Buelow (Loriot)	Herbert Grönemeyer	Nena
Harpe Kerkelingen	Peter Maffay	Nina Hagen
Stefan Raab	Udo Lindenberg	Heike Makatsch

Table 4: Names of the famous German personalities presented in the famous voice-recognition test (n = 42).

3.5.2.6 Procedure

Voice samples of the 42 famous and the 20 unfamiliar voice samples were presented in random order. After each sample, patients were asked on a written screen to categorize the voice as familiar or unfamiliar. If categorized as familiar, the task was to choose the speaker out of 3 photographs and associated names. A fourth option was a question mark symbol to provide the possibility that the patient recognized the voice as familiar but could not associate it with one of shown personalities. If the sound sample was categorized as unfamiliar, the next voice sample was presented.

At the beginning of the experiment, patients were asked to estimate their weekly exposure to television and radio (in hours). One practice trial aimed to familiarize patients with the task. In the second part of the test, we assessed each patient's familiarity with the identities of the famous voices they heard earlier and additional celebrities not being presented in the main test. Patients were shown each famous person's written name and face (in a different order than in the main test). First, patients were asked to indicate whether the famous person is familiar or unfamiliar. If the person was familiar to them, three questions followed: (i) How often have you heard the voice? (ii) How good do you think you would recognize the person's voice (iii) How good do you think you would recognize the person's face? For question (i) patients rated on a scale from 1 (never) to 5 (very frequently). For question (ii) and (iii) the answer was given on a scale from 1 (not at all) to 4 (very good). The complete experiment took about 45 minutes.

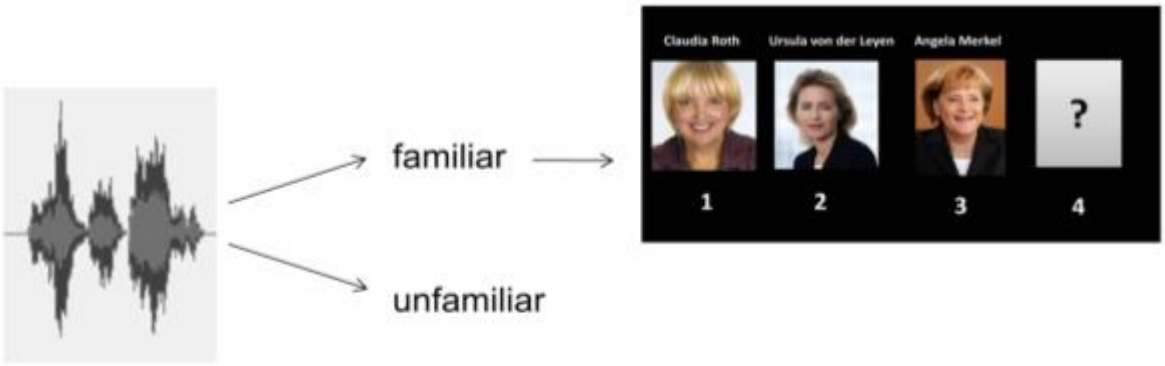


Figure 4: Famous voice recognition task: A phrase sequence (5 s) was presented via earphones. Patients chose between familiar and unfamiliar voices. Familiar voices had to be assigned to the according personality shown in one of the photographs. Option 4 could be chosen if none of the three personalities could be assigned to the heard voice.

3.5.2.7 Analysis

Based on their personal familiarity ratings obtained from the second part of the test, we individually re-classified all famous voices (n = 42) into famous and non-famous voice categories. Celebrities who were known to the patients and had been listened to at least *sometimes* (3 on the scale of question (i)) and were judged to be recognizable by voice information at least as *bad* (2 on the scale of question (ii)) were categorized as subjectively familiar persons. All other celebrities were categorized as unfamiliar for the individual patient. The non-famous voices (n = 20) were all classified as non-famous. This procedure allowed to determine whether the participants correctly or incorrectly categorized any subjective famous voice sample as familiar or unfamiliar in the first part of the test.

By applying Detection theory (Macmillan and Creelman, 2004), we computed measures of sensitivity (*d'*) and response bias on a criterion (*c*). We analyzed the data regarding two different aspects; familiarity decision making and familiar voice naming. Our analysis was based on a two by two factorial design. Assessing the *d'* and *c* value of the familiarity decision making rate, the analysis was based on the factor famous/ non-famous voice stimuli and the factor familiar/ unfamiliar response. For each participant, the proportion of correctly identified famous voices as familiar (Hit) and non-famous voices classified as familiar (False alarm) were computed. With *d'* we measured the sensitivity to differentiate between famous and non-famous voice stimuli. Patients' decision rule about the division between familiar and unfamiliar response at a specific criterion was computed with the response bias *c*.

In addition to applying detection theory to patients' familiarity decisions, we also computed proportion of those familiar voices that were correctly identified by their names. Patients listened to one voice stimulus and then chose the according personality out of three possible personalities that appeared on the screen as photographs with names. One fourth option was a question mark that patients could choose if they had recognized the voice as familiar but could not associate the voice with any of the presented photographs.

3.5.3 Tests for investigating the Specificity of Voice Recognition Impairments

To investigate whether potential voice processing impairment in patients was accompanied by a face recognition deficit (Van Lancker and Canter, 1982; Luzzi et al., 2017b) or were specific to voices (Neuner and Schweinberger, 2000), we tested participants' abilities in face processing.

We used a standard (CFMT) and a novel face processing test.

3.5.3.1 Face-Name Learning

We developed a face-name learning test to assess participants' abilities in associating faces with given names. The participants were exposed to six unfamiliar male faces in 4 learning phases. A learning phase was always followed by a testing phase except for the "final test" that started automatically after testing phase 4. Each face appeared in different views for 6 times in the first learning phase and 2 times in the following 3 learning phases. The name was shown together with the photograph. Participants were instructed to memorize the right combination of face and name. In the testing phase, we assessed how well the participants had associated face and name by asking the participants to choose the right name out of the 6 presented names after the appearance of one photograph. The first two learning phases were noise-free while learning phase 3 and 4 were added with increasing noise.

3.5.3.2 Stimuli and Presentation Software

The stimulus set consisted of six pictures of male faces comprising three British and three Spanish actors. The pictures were downloaded from freely accessible websites and edited with Adobe Photoshop CS4 (version 11.0.2; Adobe Systems, Inc., San Jose, CA, US). Thirty-eight versions of each picture were generated by adding different levels of Gaussian noise (e.g., 15%, 30%, 60%) yielding a total of 228 images. Original photographs were taken from different views with varying facial expressions and under changing lightning conditions. Thus face images varied in a natural extent. The experiment was implemented in the Presentation software (Neurobehavioural Systems, Inc., CA, USA).

3.5.3.3 Procedure

The experiment was divided into four learning and five test blocks. During the learning phase, the participants were exposed to the six face-name associations (common German first names were chosen; e.g., Peter, Jan, Timo, Alex, Otto, Leon). During the first block, two sequences of five face-name pairings per person were shown. During the second block, three face-name pairings per person and during the third and fourth block two face-name parings per person were provided in randomized order. In the test blocks, the participants were asked to carry out a six-alternative forced choice task which assessed participants' learning performance. The participants were presented with a picture of one of the learned individuals and were then prompted to select the target name. Five items per person were tested in each test block. To prevent ceiling performance in the second, third and fourth test blocks, images were presented with increasing noise levels. In the final test block, the images without additional noise were presented again. The participants were given feedback on both correct and incorrect responses. In addition, the correct combination of face and name was repeated after feedback during the first and the second test blocks. Response was recorded via keyboard. The experiment took approximately 30 minutes.

3.5.3.4 Cambridge Face Memory Test (CFMT)

Using the CFMT (Duchaine and Nakayama, 2006), we investigated the recognition abilities of six learned male faces. After having learned six male faces participants were asked to recognize one learned target face out of 3 presented faces. Face recognition was tested in three different test sections: same images, novel images and novel images added with Gaussian noise. There is a total of 72 items. We applied the upright version of the test.

3.5.4 Tests for impairments in processing basic acoustic voice properties

Voice as a complex acoustic construct is composed of temporal-spectral properties such as fundamental frequency (i.e., pitch) and spectral formant frequencies (i.e., voice timbre). These properties were characterized as having a major perceptual relevance to identifying familiar speakers (Lavner et al., n.d.; Smith et al., 2005; von Kriegstein et al., 2006). The pitch and timbre tests served to investigate whether voice recognition impairments are associated with deficits in perceiving these basic acoustic properties.

3.5.4.1 Stimuli and presentation software

The stimulus set consisted of five English vowels (/a/, /e/, /i/, /o/, /u/) resynthesized using the STRAIGHT software package (Kawahara et al., 2008) implemented in Matlab (version 7.7, The MathWorks, Inc., MA, USA). The original vowels were spoken by a male speaker (same material used in (Smith et al., 2005)). Prior to being resynthesized, the vowels were modified to be monotonic and 600 ms in duration. For the pitch-discrimination task, all tokens of a given vowel were identical except for their fundamental frequency. F0 is the physical correlate of a speaker's glottal-pulse rate (GPR), which determines their voice pitch. For the timbre-discrimination task, all tokens of a given vowel were identical except for their spectral envelopes, which were scaled proportionally up or down in log-frequency space from the original spectral envelope. Spectral envelope is the physical correlate of a speaker's vocal-tract length (VTL), which is an aspect of vocal timbre that correlates very strongly with overall speaker size (Fitch, 1997). Both tests were implemented in Python (version 2.7.3, http://python.org/) and responses were recorded via a keyboard.

3.5.4.2 Procedure

We used an adaptive-tracking procedure (Kaernbach, 1991) to measure the participants' pitch and timbre just-noticeable differences (JNDs). During the pitch-discrimination test, participants listened to pairs of sequentially presented vowels differing only in their F0. One vowel per trial always had an F0 of 112 Hz, and the other was higher in F0 by an amount (ΔF0) defined in musical cents (1 semitone = 100 cents). The order of the stimuli was random on each trial, and participants reported which one was higher in pitch. The initial ΔF0 was 100 cents; this value decreased in steps of 10 cents following each correct response and increased in steps of 30 cents following each incorrect response. After four reversals (a switch from correct to incorrect response or vice versa within two consecutive trials), the up and down step sizes were changed to 6 and 2 cents, respectively, and the block of trials continued for a further 10 reversals. JND was estimated from a single run by taking mean of all ΔF0 values visited during the final 10 reversals, and the participant's overall JND was defined as the mean JND over five runs. Feedback about response accuracy followed each trial. At the beginning of the test,

participants were familiarized with the auditory stimuli by presenting them with two vowels at the extremes of the F0 range. The average test duration was 15 minutes.
The timbre-discrimination test was identical to the pitch-discrimination test except that the stimuli on each trial differed in their spectral envelopes, and participants reported which vowel was spoken by the smaller speaker. One vowel on each trial had a spectral envelope equal to that of the original speaker, while other differed by ΔSER, defined in percent. Initial ΔSER was 12%; up and down step sizes were 3% and 1% for the first four reversals, and 0.6% and 0.2% for the remaining 10 reversals.

3.6 Statistical Analysis

Behavioral data were analyzed in IBM SPSS 22. We performed two different types of analyses. In a first step, we conducted a group comparison between right-and left-hemisphere damaged patients. We calculated mean scores and standard deviations for each subtest for the entire patient group, left-hemisphere damaged patients and right-hemisphere damaged patients. For each subtest, we conducted group comparisons between right-and left-hemispheric patients using independent sample t-tests.

Pearson correlations were used to assess whether the test performance was influenced by demographic values (age, time since onset, education), hearing level, or neuropsychological performance. Moreover, we aimed to assess influence of acoustical vocal features on the voice processing performance. We therefore computed correlations between the voice recognition subtests and auditory processing (see supplementary material) with bivariate correlation analysis.

In a second step of analysis, we derived a patient group consisting of patients that performed below the 25th percentile of the Median on the voice-learning tests (P25= 50.31%). This group will be referred to as voice deficit patients (VD patients) in the following. For all tests, we compared test performance of the VD patient group to the remaining patient sample, i.e. other patient group (O patients) by employing the Welch`s Test. The Welch`s test was chosen due to unequal variances and unequal sample size of VD and O patients. We assessed the inequality of variances between these groups with the Levene’s test.

We computed performance differences within subjects with a one-way repeated measures ANOVA. We investigated performance differences on the voice-name and voice-face test within subjects to reveal whether voices were learned easier in combination with a face or a name. We were further interested whether performance in these tests was different between female and male voices (dependent variable gender).

We analyzed structural MRI data with voxel lesion symptom mapping (VLSM) (Bates et al., 2003b). We performed VLSM to identify any systematic relationship between damaged voxels and behavioral measures by using VLSM 2.55 (http://www.neuroling.arizona.edu/resources.html) implemented in Matlab (version 8.2, The MathWorks, Inc., MA, USA). For details on the VLSM analysis please refer to the attached paper *'Voice-identity recognition deficits are induced by lesions in the temporal and inferior parietal lobe'* in chapter 13 (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K., in prep.).

4 RESULTS

4.1 Neuropsychological Data

All patients performed above the cut-off value on the neuropsychological assessment. The cut-off value for the neuropsychological assessment was the 2nd percentile rank (PR) for each test. Results for neuropsychological tests are presented in Table 5.

In the TAP (test of attentional performance) 1 patient performed below the 2nd percentile rank in the intrinsic attention condition but above the 2nd percentile rank in phasic attention condition. The mean value was 39.67 PR (*SD= 28.41*) for the intrinsic condition and 31.79 PR (*SD= 24.70*) for the phasic condition.

For the digit and spatial span test we computed a composite score, i.e. mean of backward and forward condition for both subtests ("working memory score"). The mean working memory score for the entire group was 35.11 percentile rank (*SD* = 19,29). 1 patient scored below the cut-off in "spatial span backward" but above cut-off on all other tasks.

The mean score on the "Wortschatztest" was 53.02 percent (*SD* = 21,32). No patient performed below the cut-off.

Two scores were taken into consideration for the "Gesichter-Namen-Lerntest": Percentile rank for the sum of correctly learned names of the 4 testing phases and the percentile rank of cued recall of names after 30 minutes (free recall of names remembered) and a memory score. The mean cued recall was 76.43 percent (*36.22*). One patient scored below the 2nd percentile rank.

Among all neuropsychological tests, all patients scored above the cut-off value in at least 8 of the 10 subtests. We therefore did not exclude patients due to their neuropsychological performance.

Next, we tested whether performances in the NPS different between the right- and left hemisphere damaged patient group. The group comparison revealed comparable performances in both patient groups.

We calculated a composite score over all neuropsychological tests, i.e., mean percentile ranks of TAP, GNL, working memory and WST. There was no correlation between

neuropsychological performance and the performance on the person recognition test battery.

Table 5 sums up neuropsychological performance of the patient sample.

Test (PR)		**Mean (*SD*)**	
	All Patients	**RH**	**LH**
Vocabulary test (WST)	53.42 (*21.46*)	53.33 (*18.84*)	53.56 (*25.99*)
Working memory			
Digit Span	30.39 (*18.87*)	34.42 *(20.99*)	23.78 (*12.86*)
Spatial Span	39.89 (*23.95*)	44.63 (*24.46*)	32.11 (*21.71*)
TAP			
intrinsic	39.67 (*28.41*)	39.62 (*33.08*)	39.77 (*19.90*)
phasic	31.79 (*24.70*)	31.09 (*29.12*)	32.92 (*16.21*)
Face-Name Learning (GNL)			
GNL cued recall	76.43 (*36.22*)	81.69 (*32.95*)	67.78 (40.83)
GNL Sum D1-D4	56.81 (*36.71*)	64.83 (*32.95*)	43.64 (*39.11*)

Table 5: Results of neuropsychological assessment for all patients, Right Hemisphere (RH) and Left Hemisphere (LH), SD= 1 standard deviation.

4.2 Peripheral Hearing Assessment

34 patients obtained mean hearing thresholds within the normal range of 0-26 dB (mean 14.71 dB, *SD* = 7.39) on pure tone audiometry. 5 patients obtained levels above 26 dB. All of them were still within the range of *mild hearing loss* (< 40dB, (Pascolini and Smith, 2009)). Individual adjustments of the sound volume before testing allowed for normal sound perception and assured unaffected performance in our tests. There was no correlation between hearing levels and behavioral test performance. No patient was excluded due to mild hearing loss. One patient is supplied with hearing aid but did not wear them during audiogram and testing sessions. When being asked about the sound perception, he reported unimpaired hearing of test sounds during the experiment.

4.3 Questionnaire

A questionnaire assessing subjective person recognition skills was part of our comprehensive test-battery. 25 patients had neither noticed a difference between pre- and post-incident in their person recognition abilities in general, nor their voice and face recognition. The remaining 12 patients who reported generally decreased person recognition abilities (i.e. shift from *very good* to *good*, from *good* to *sufficient* or from *sufficient* to *poor*) were notably right-brain damaged (8R /4 L).

	Person recognition		**Voice recognition**		**Face recognition**	
	Before	After	Before	After	Before	After
Very good (n=37)	17	14	9	5	17	8
RH (n= 23)	9	8	5	2	11	4
LH (n=14)	8	6	4	3	6	4
Good	18	11	21	18	17	18
RH	13	7	13	11	11	12
LH	5	4	8	7	6	6
Sufficient	1	10	7	10	3	9
RH	1	7	5	7	1	6
LH	0	3	2	3	2	3
Poor	1	2	0	4	0	2
RH	0	1	0	3	0	1
LH	1	1	0	1	0	1

Table 6: Questionnaire results on subjective person recognition abilities. Number of patients who rated their abilities as very good, good, sufficient or poor, Before = time before lesion onset, After = time since lesion onset, RH= Right Hemisphere, LH= Left Hemisphere.

4.4 Behavioral Results

4.4.1 Voice-Identity Processing Skills

4.4.1.1 Voice Learning of Unfamiliar Voices

All patients completed the two voice- earning tests: the voice-name test and the voice-face test. All patients performed above chance level (33.33%). The Mean overall voice-name learning performance was 55.90 % (*SD* = 11.52), the Mean overall voice-face performance was 63.71 % (*SD* = 11.10) and the patient sample reached a Mean overall voice-learning score of 59.80 % (*SD* = 9.75). Figure 6 shows overall performance and separate test performance for the respective unfamiliar voice-learning tests. Overall scores were calculated from the subtests: Overall Voice-Name = M*ean* Voice-Name female + *Mean* Voice-Name male, Overall Voice-Face = *M* Voice-Face female + *M* Voice-Face male, Overall Voice learning = *M* Overall Voice-Name + *M* Overall Voice-Face.

Male voices were learned easier than female voices in both tests (voice-name: *p* Wilks' lamda = < 0.001, observed power: 1.0, voice-face *p* Wilks' lamda = < .001, observed power: .953) which confirmed self-reports by the patients. The results show that patients also performed better on the voice-face than the voice-name tests (*p* Wilks' lambda = < 0.001, observed power: .978).

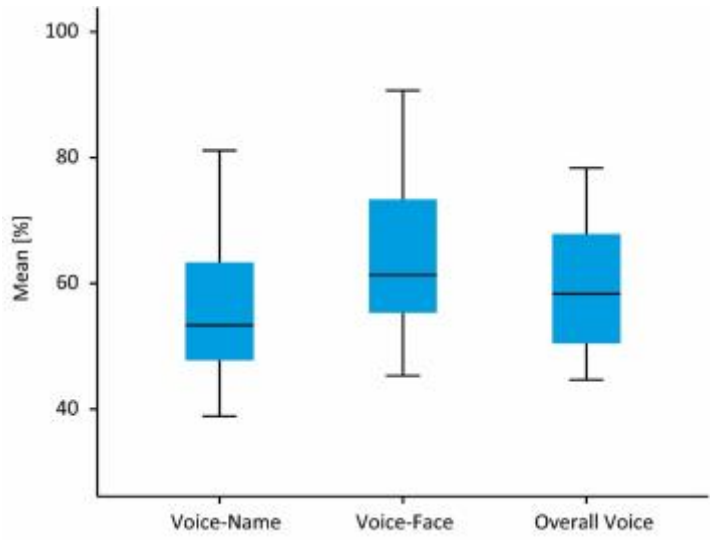


Figure 5: Performance on the voice-learning tasks for the entire patient sample. Voice-Name learning, Voice-Face learning, Overall Voice performance (%). Error bars represent 1 SD.

4.4.1.2 Voice Recognition using Famous Voices

36 patients completed the "famous voice recognition test". One patient did not participate in this test because her mother tongue was not German and we assumed that her exposure to German media in the past 11 years was not sufficient to be comparable to native German speakers.

We calculated measures of sensitivity to distinguish between famous and non-famous voices (*d'*) and patients' decision rule about the division between familiar and unfamiliar response (*c*). The mean d prime for the entire patient sample was 1.52 (*SD = 1.30*). The mean c value for all patients was -2.7 *(SD = .59).* On the subtest for naming (semantic association) the group Mean was 70.92 % *(SD= 17.49).*

4.4.2 Processing Basic Acoustic Properties: Pitch and Timbre

We determined patients' GPR and VTL discrimination thresholds for the five vowels (a/e/i/o/u). The lower the threshold in JNDs (just noticeable differences in cents) the better the ability to discriminate vocal pitch sounds. Group Mean for JND on the Pitch test was 124.14 cent *(SD= 63.40).* On the timbre test, the group Mean for JND was 9.39 (SD = 4.48) SER (spatial envelope ratio).

4.4.3 Tests for investigating the Specificity of Voice Recognition Impairments

4.4.3.1 Face Recognition Performance

All patients performed above chance level on the Face-Name test. Mean score for the patient sample was 57.92 (*SD* = 19.67) on the Face-Name test.

On the Cambridge Face Memory Test, the mean score of the entire patient group was 67.20 % (*SD* = 14.27). The cut-off value for prosopagnosia would be 2 SD below the control mean of a previous study (Duchaine and Nakayama, 2006). 9 patients performed below this score and were consequently indicative of prosopagnosia. 8 out of these 9 patients presented right-hemispheric damage.

4.4.4 Correlation between Acoustic Properties and Voice Learning

We found positive correlations between the two measured basic acoustic properties (i.e. Pitch and Timbre) and the overall voice-learning performance of all participants. Better performance on the Pitch and Timbre task (lower JND) was correlated with better performance on the voice-learning tasks. Correlations are presented in in Figure 6 (Overall voice_timbre: *r*= -.596**, *p* = < .01, voice_pitch r= .365*, *p* = .026). See supplemental material for details.

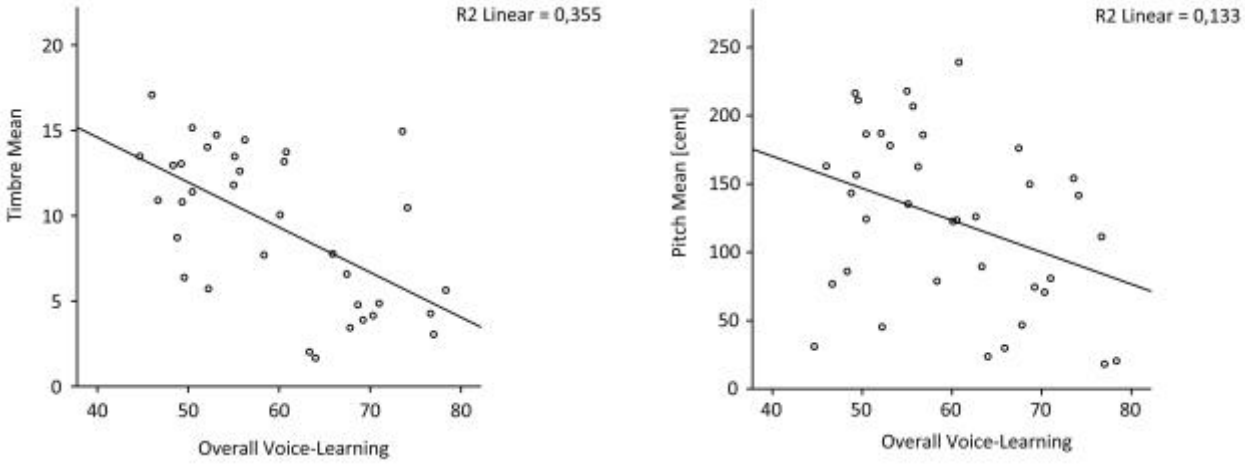


Figure 6: Left: Correlation of Voice-Learning and Timbre as a basic auditory cue: r= -.596 **, p= < 0.01. Right: Correlation of Voice-Learning and Pitch as a basic auditory cue: r= .365*, p= .026

4.5 Performance Comparison of Right-and Left- Hemispheric lesioned Patients

According to our hypothesis, we expected patients with unilateral lesions in the right hemisphere to show lower performance on the voice and face tests than patients with unilateral lesions in the left hemisphere. This hypothesis could not be confirmed through statistical analysis in the given sample. Group differences in the performance between right- and left- hemisphere on the behavioral tests were not significant. However, there was a tendency for lower task performance in especially the voice-learning tests in the right-brain damaged group (*p* overall voice= .116, *t* overall voice= 1.61). A larger variance on the voice-learning tests in the right-brain damaged group supported the hypothesis that impaired voice recognition abilities cannot be explained through general right-hemispheric damage but are likely due to lesions in specific subregions in the right

hemispheric areas. Figure 7 depicts performance of right- and left-hemisphere lesioned patients on the three voice-learning test modalities.

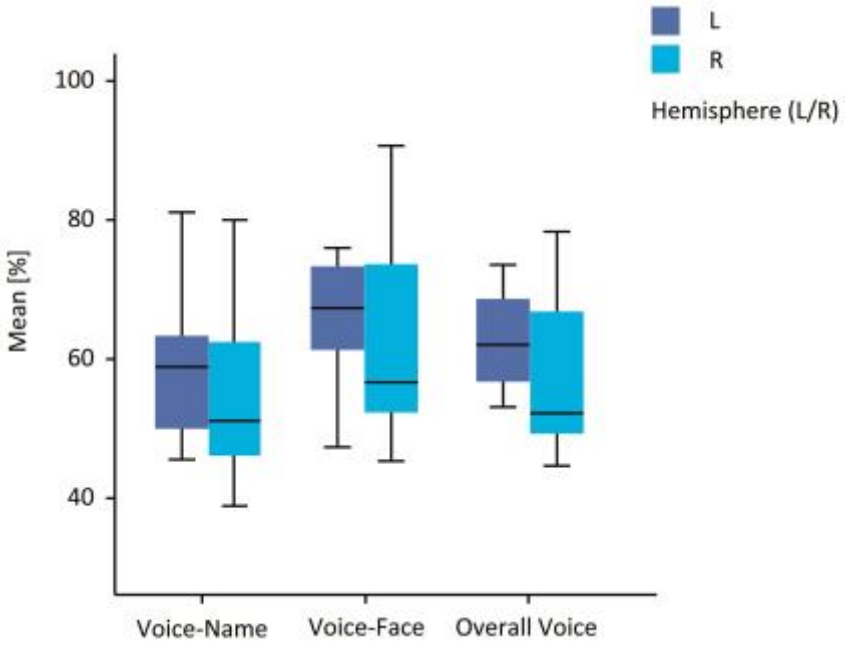


Figure 7: Performance comparison of right- and left hemisphere lesioned patients on the voice-learning tests. Larger variance is present in RH compared to LH. Error bars represent +/ - 1 SD.

4.6 Patients presenting Voice-Identity Processing Impairment

To characterize the patterns of possible voice recognition deficits in this sample, we employed a second step of analysis. Here, we investigated a group of patients that scored below a cutoff value (25th percentile rank of the median) in the overall voice learning performance (*Mean* of voice-name and voice-face test = 59.71, *Median*= 58.34, *25th Percentile*= 50.45) This group is called VD patients (Voice Deficit). We compared VD patients to the remaining other patients of the sample (OP) to characterize the specific performance pattern in the VD patients, i.e. selective voice recognition deficits or general person recognition deficit by voice and face.

VD patients comprised nine patients. One notable fact was that all VD patients had right-hemispheric damage which is in accordance with our hypothesis. Mean age was 47.00 years (*SD* = 12.97) which did not differ significantly from the rest of the patient group. Mean recovery time of these 9 patients was 31.20 months (*SD* = 15.51).

In our person-recognition questionnaire, three patients in the VD group (RH) reported a decrease in their voice recognition abilities since brain injury. The remaining patients rated their abilities as either *very good* or *good*, before and after brain injury.

We further confirmed our findings of weak performance on the voice tests by performing statistical group comparisons (p *overall* = 0.00, *Voice-Name* = 0.003; p *Voice-Face* = 0.01) between VD and O patients on the voice-learning tasks.

Table 7 and Figure 8 show results of VD and O patients on the voice-learning tests.

	Mean *(SD)* VD patients	**Mean *(SD)* O patients**	***p* *(Welch`s Test)***
Voice-Name	44.81 *(4.90)*	59.46 *(10.75)*	< .001
Voice-Face	51.40 *(5.66)*	67.43 *(9.77)*	< .001
Overall Voice-Learning	48.11 *(1,90)*	63.45 *(8.35)*	< .001

Table 7: Voice-Learning Performance Comparison between VD patients and O patients. Significance level p= < 0.05.

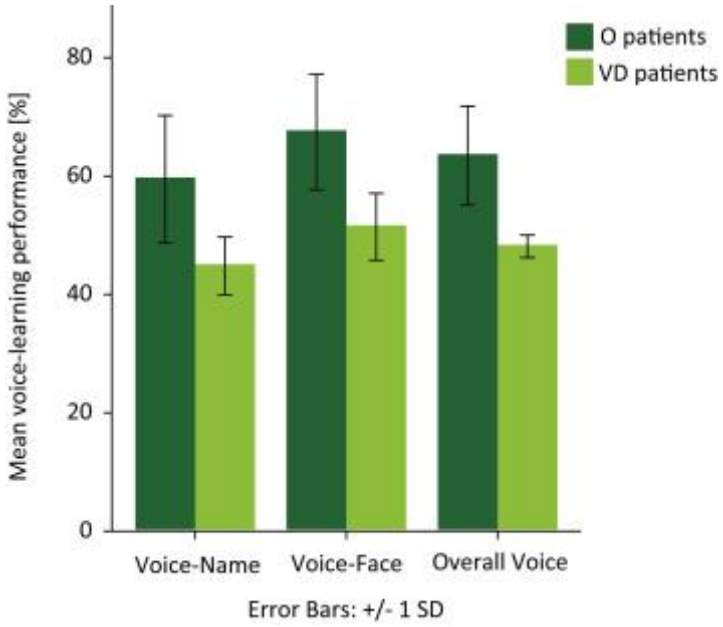


Figure 8: Group comparison of voice-learning performance for VD patients and O patients (p= <0.01) Error bars represent +/ - 1 SD.

We investigated mean differences between control modalities (e.g. face tests and neuropsychology) and voice tests to determine whether poor voice-learning performance of VD patients was a specific deficit. The distinction between the mean differences of the two test modalities was not significant due to the large variance within the patient groups (p = 0.9). For details see table 8.

VD patients performed similar to O patients when distinguishing between familiar and unfamiliar voices (Mean d value VD: 1.82, *SD* = 1.88 Mean O: 1.42, *SD*= 1.11, p = 0.59). The response bias (c) for VD patients was -.193 (*SD* = .488). The mean c value for O patients was -.276 (*SD* = .596), p = .68)

Beside the familiarity decision task, the test on familiar voice recognition comprised naming, i.e. semantic association to voices that had been recognized as familiar. While O patients named 72.85 % (*SD* = 14.83) of the familiar personalities correctly, VD patients obtained 64.17 % (*SD* = 24.77, p = .371).

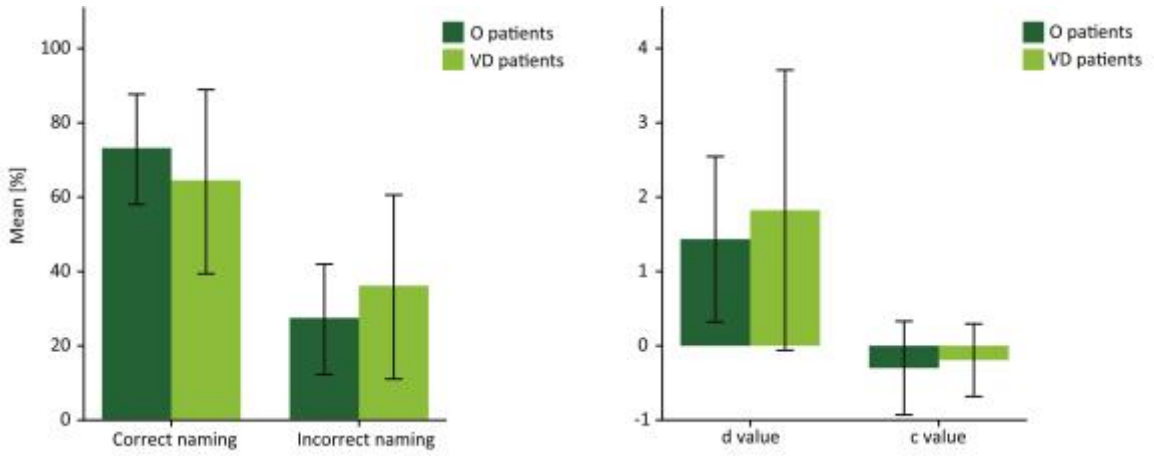


Figure 9: Performance of O and VD patients on the famous voice recognition test. On the left: Correct and incorrect naming in Mean [%]. Error bars represent +/- 1 SD. On the right: d' and c values of VD patients and O patients on familiarity decision making.

VD patients reached a mean score of 141.24 cent (*SD* = 21.29) on the pitch task (GPR). Thresholds for the single vowels ranged between 18.28 and 239.08 cent and the mean score for O patients was 122.32 cent (*SD* = 64.34). The task had been described as

difficult by several patients. The mean SER of VD patients was 12.06 SER (*SD = 3.26*) on the timbre task (VTL).

VD patients reached a mean score of 53.19% (*SD* = 14.50) on the Face-Name test. The mean score of VD patients on the CFMT was 58.79% (SD: 19.05). For details see table 8.

Test	**VD patients *(SD)***	**O patients *(SD)***	**F**	**Df1**	**p-value**
Neuropsychology (composite score)	50.02 *(12.65)*	47.20 *(14.65)*	.262	32	.590
Voice Tests					
Voice-Learning	48.11 *(1.90)*	63.44 *(8.35)*	29.29	35	<0.001**
Famous Voice Recognition					
Familiarity judgment (d`)	1.82 *(1.88)*	1.43 *(1.11)*	.554	34	.591
c	-.19 *(.488)*	-.30 *(.63)*	.195	34	.618
Naming/ Semantic association (%)	64.17 *(24.77)*	72.85 *(14.83)*	1.558	34	.371
Face tests	55,99 *(15.04)*	63.47 *(14.77)*	1.731	35	.215
Acoustical voice tests					
Pitch (cent)	141.23 *(63.89)*	118.64 *(63.41)*	.861	35	0.37
Timbre (SER)	12.05 *(3.25)*	8.47 *(4.52)*	4.736	33	0,019*

Table 8: Group Comparison on all tests between VD and O patients, (SD) = 1 Standard Deviation, significance level p < 0.05

4.7 Single Case Reports

To further investigate possible selective voice-identity processing deficits within the VD group, we correlated their overall voice-learning performance with the control modalities (neuropsychology and face tests). This allowed for a visual depiction of performance on the voice and control tests. Figure 10 plots the correlation between overall voice- learning performance and the control modalities for all patients from our sample. Here, we discovered that three patients from the VD group appeared to have a selective deficit in voice processing, i.e. low performance in voice-learning tests and high scores on the control tests.

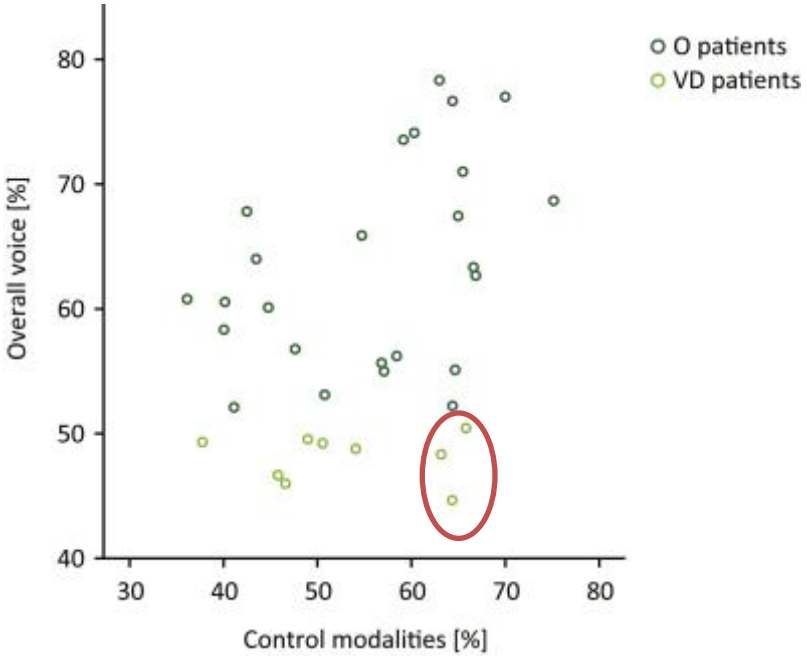


Figure 10: 3 patients with selective voice processing deficits. This figure plots interaction of overall voice-learning and control modalities (face tests and neuropsychology). Light green: VD patients, dark green: O patients. 3 patients with high scores on control tests and low scores on voice-learning tasks are highlighted (red circle).

Table 9 displays individual test results of VD patients as well as their lesion location. 3 patients with voice-selective performance deficits are highlighted. We describe their individual impairment pattern along with their individual lesion location in the following.

Patient	**1**	**2**	**3 ***	**4**	**5**	**6**	**7**	**8 ***	**9 ***
Hemis-phere	R	R	R	R	R	R	R	R	R
Lesion Location	Put pCI STG	POG FFG ITG SFG Thal Hipp Put	SMG STG MOG AnG	sFG OS aIG Put mFG STG peCS RO	PS mOG STG STS RS peCS poCS iFG Put AnG	LOG AnG PS FFG	STG STP iPS peCG AnG aIG	pMTG	CoI STG Put
Test Mean									
Voice Tests [%]									
Voice-Name	51.12	40.00	40.00	46.67	52.23	47.78	43.34	38.89	43.33
Voice-Face	47.33	58.67	56.00	46.67	45.34	51.34	48.67	62.00	53.33
Overall Voice	49.23	49.33	48.00	46.67	48.78	49.56	46.00	50.45	48.33
Famous Voices									
c	-.246	-.468	.453	.308	.082	-.777	.000	-	-.877
d’	.738	.593	.852	4.536	.618	1.325	5.152	-	.809
Naming	43.24	55.56	33.33	89.74	36.84	90.00	89.65	-	75.00
Face Tests [%]									
Face-Name	53.33	30.67	41.33	48.00	49.33	44.67	74.67	68.67	68.00
CFMT	59.72	34.72	69.44	43.06	40.28	55.56	54.17	93.06	79.17
Overall Face	56.53	32.70	55.39	45.53	44.80	50.12	64.42	80.87	73.58
Auditory Tests									
Pitch [cent]	216.44	156.56	31.04	76.84	143.31	211.20	163.10	186.60	86.06
Timbre [SER]	10.81	10.80	13.48	15.38	8.72	6.38	17.08	15.15	12.95

Table 9: Individual results from VD patients on all behavioral tests and their lesion location, *) marks three patients with selective voice-learning deficits. aIG= anterior Insula gyrus, AnG= angular gyrus, CoI= Cortex of Insula, FFG= Fusiform Gyrus, Hipp= Hippocampus, iFG= inferior frontal gyrus, iPS= intraparietal sulcus, ITG= Inferior temporal gyrus, LOG= lateral occipital gyrus, mFG= midfrontal gyrus, MOG= midoccipital gyrus, OS= Orbital sulcus, pCI= posterior Cortex of Insula, peCS= precentral gyrus, pMTG= midtemporal gyrus, poCS= postcentral sulcus, POG= posterior occipital gyrus, posterior division, PS= parietal sulcus, Put= Putamen, RO= Rolandic operculum, RS= Rolando sulcus, SFG= superior frontal gyrus, SMG= superior marginal gyrus, STG= superior temporal gyrus, STP= superior temporal pole, STS= superior temporal sulcus, Thal= Thalamus. Anatomical location was defined using neuroanatomical atlases in MRicron (Rorden and Brett, 2000).

Patient 3 is a 58-year-old male patient after ischemic stroke 42 months prior to testing in the right temporal and parietal and occipital lobe. Precise lesion location was the superior marginal and superior temporal gyrus, the angular gyrus and mid occipital gyrus. On the questionnaire, he reported his voice recognition abilities had shifted from *average* to *poor* after brain injury, while his face recognition abilities had changed from *very good* to *average*. This patient performed below group mean on the voice-learning test (total score of 48.00 %) and below the entire group mean on the famous voice recognition task (sensitivity (d`): .852 and naming: 33.33%).

His neuropsychological and face-learning scores were above average of VD patients (NPS; 73.31%, Face: 55.39 %). Patient 3 further performed above the group mean on the control tests for vocal pitch discrimination (Mean VD patients: 127.11 cents, Mean patient 3: 31.04 cent). For vocal timbre discrimination, his score was close to the average VD patients' mean (Patient 3 mean: 13.48 SER, VD patients mean: 12.06 SER).

Patient 8 is a 32-year-old female who was diagnosed with oligoastrocytoma in the right temporal lobe, specifically the right posterior division of the mid temporal gyrus. Tumor extraction had been performed 12 months before the experimental investigation. This patient is originally from Russia but has lived in Germany for 11 years and is fluent in German. She reported that since tumor extraction she has severe difficulties in recognizing familiar voices, e.g. her partner's voice or familiar people's voices on the phone if they do not mention their name. On the questionnaire, she rated her voice recognition abilities as *poor* after brain injury relative to *good* before tumor extraction, and noted a shift in face recognition from *very good* to *good* after tumor extraction.

As mentioned above, this patient did not complete the famous voice recognition test due to her fewer contact to German media compared to the other patients. Her performance on the voice-learning task was assumed to not be affected by a language acquisition later in life as there is evidence that voice recognition is well preserved for a non-native language (Fleming et al., 2014). This patient was fluent in German and German was the spoken language with her partner and family at home. However, this patient performed below the 25th percentile on the voice-learning tasks whereas she performed above group mean on the face tests (Patient 8 mean: 80.87 %, Mean VD patients: 54.33% (*SD* = 15.13)). On the neuropsychological test battery, her performance

was comparable to the Group Mean for VD patients (Patient 8: 50.78 %, VD patients: 50.02 %, *SD*: 14.02).

She reported difficulties in performing the auditory control task for pitch and timbre, i.e. recognizing differences between two presented sounds as they sounded equal to her in all conditions. Her performance on this auditory test battery was below the mean of VD patients (patient 8: 186.6 cent, 15.15 SER, VD patients Mean: 141.24 cent (*SD* = 21.29), 12.06 SER (*SD = 3.26*)).

Patient 9 is a 57-year-old female patient who experienced ischemic stroke 13 months prior to the current investigation. Her lesion is located in the right hemisphere, along the insula cortex and the superior temporal gyrus, as well as the Putamen. Unlike the other 2 described patients this patient had not noted any difficulty in recognizing familiar people by their face or voice. She rated her voice and face recognition abilities as *good* before and after brain damage on the questionnaire.

Her overall performance on the voice-learning task was 48.33 %. This patient scored 73.58 % (above Group Mean) on the face tests and 52.75% on the neuropsychological test battery (Mean VD patients: 50.02, *SD* = 14.02). On the auditory tests her performance was above Group Mean of VD patients (Pitch: 86.06 cents, Timbre: 12.95, for references see paragraph above).

For an overview of the lesion locations of the three single cases please refer to Figure 11.

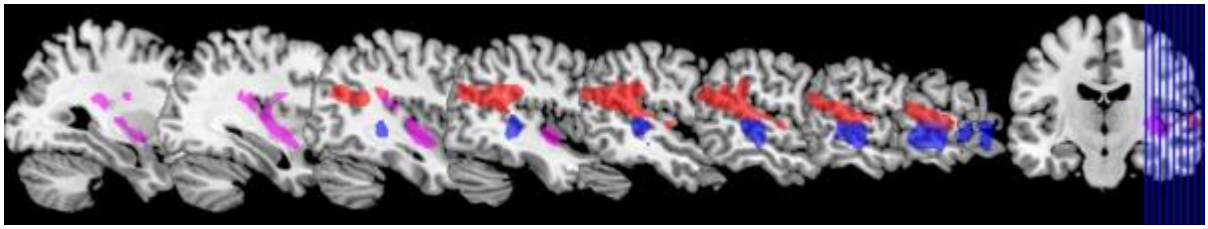


Figure 11: Lesion overlay of 3 patients with selective voice processing deficits. Red: Patient 3, SMG; STG, MOG, anG, blue: Patient 8, pMTG, violet: Patient 9, CoI, STG, Put (see Table 9 for details).

4.8 Outlook: Neuroanatomical Analysis using VLSM

Next to the behavioral analysis, neuroanatomical correlates of voice recognition deficits were of great interest to our research. To this end, a voxel-based lesion symptom mapping (VLSM) analysis was conducted (Bates et al., 2003a). The analysis was based on the current patient sample and the collected behavioral data, yet extended to a larger sample size (n= 58). Results of the VLSM analysis are in preparation for publication (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein, in prep.) and are attached to the current work in chapter 13.

VLSM has successfully been applied across domains of cognition throughout the past decade, i.e. speech production and comprehension (Dronkers et al., 2004; Borovsky et al., 2007) as well as aphasia (Henseler et al., 2014). VLSM allows for a voxel-wise statistical correlation between continuous behavioral measures and corresponding brain lesions. Preselecting patients based on their lesion locations or symptom complexes as in other methods is not required in VLSM. All lesioned voxels and behavioral measures are considered in the analysis which results in precise information on connections between lesioned brain areas and deficits in certain tasks, e.g. voice-learning and processing tests. VLSM analyses run a general linear model comparing performances on every measure in patients considering lesion status of the respective voxel (0= intact, 1= lesioned).

The conducted VLSM analysis revealed two key findings on brain regions associated with voice recognition impairment. First, voice recognition impairments were associated with lesions in the temporal and the right inferior parietal lobe. When controlled for face recognition deficits, the association for voice recognition deficits remained in the right temporal lobe only.

The second finding is an association of lesions in the right inferior parietal lobe with voice recognition deficits if the voices had been learned together with a face. In contrast, deficient voice recognition after voices had been learned together with a name was associated with right temporal lobe lesions. This suggests that the inferior parietal lobe might be involved in multimodal person recognition when voices and faces are integrated. Key findings are depicted in Figure 12 and 13.

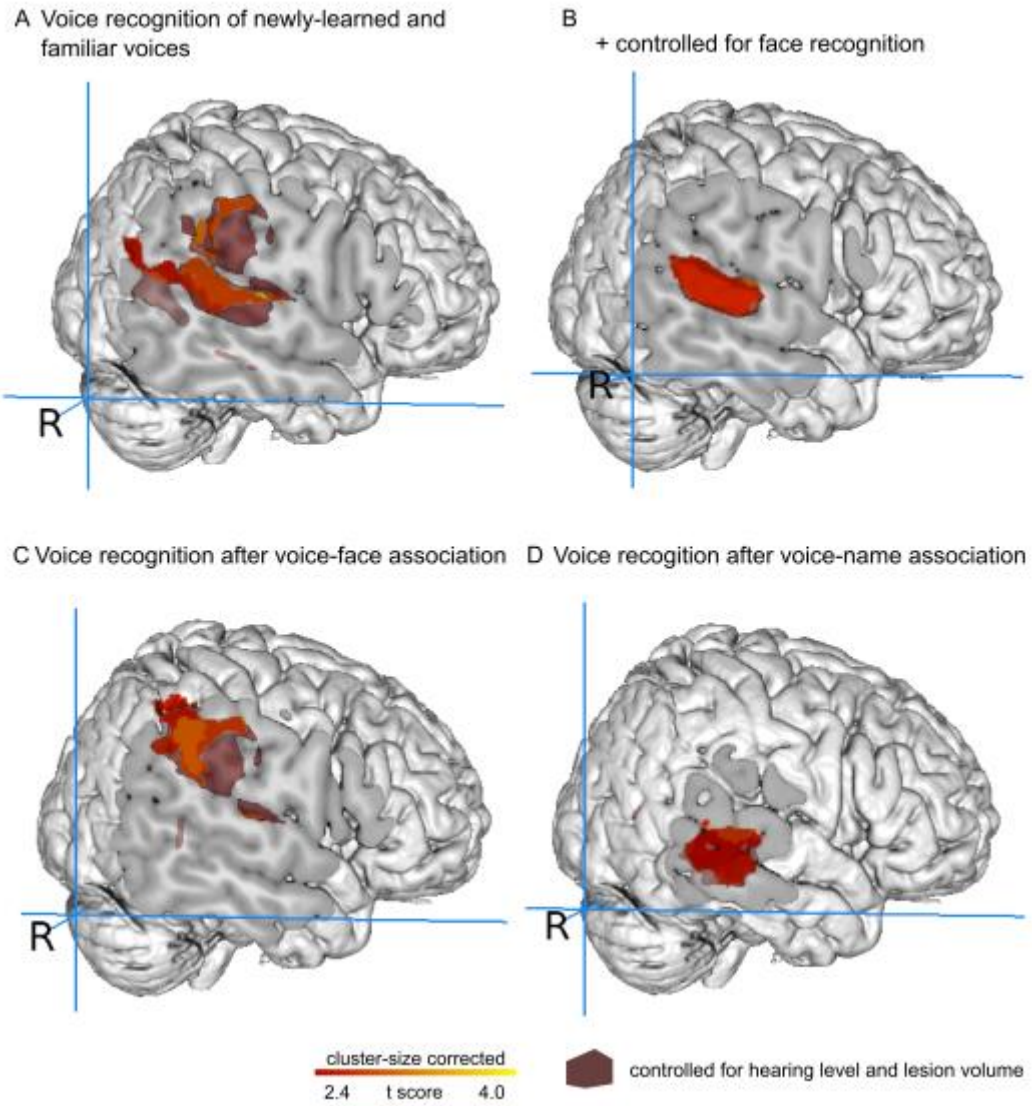


Figure 12: A. VLSM results for newly-learned voice recognition (composite score of voice-name and voice-face test, *n* = 43 patients). **B.** VLSM results for newly-learned voice recognition controlled for face recognition (i.e. composite score of CFMT and face-name test, *n* = 40). **C.** VLSM results of the voice-name test (*n* = 58). **D.** Results of the voice-face test (*n* = 43). VLSM results that are controlled for hearing level and lesion size are overlaid as red surfaces on each image (except B as results of this analysis reached no significance). Analyses were restricted to the bilateral temporal lobe and the right inferior parietal lobe. All voxels shown exceeded the critical threshold for significance ($p < 0.01$ cluster-size corrected, 1000 permutations) (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K., in prep.)

Moreover, VLSM analysis discovered distinct brain regions for voice identity recognition of different familiarities. Deficits in recognizing newly learned voices were associated with right-hemisphere lesions while familiar voice recognition deficits were associated with lesions in the left hemisphere, specifically the posterior MTG.

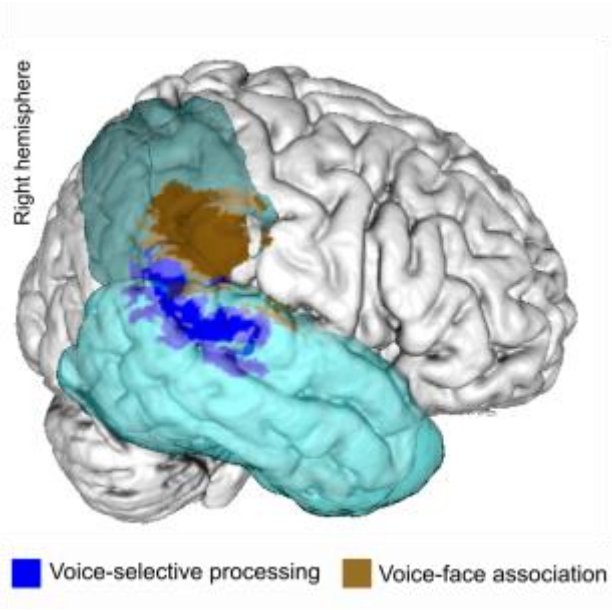


Figure 13: Overview of lesions associated with voice-recognition deficits within the temporal (cyan map) and inferior parietal lobe (turquoise map) (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K., in prep.)

Taken together, this study provides important insight into the neuronal mechanisms of voice identity recognition. It highlights the crucial role of the right temporal lobe in the recognition of voices as well as the involvement of the right inferior parietal lobe in voice identity recognition.

For the complete article please see the attached paper '*Voice-identity recognition deficits are induced by lesions in the temporal and inferior parietal lobe'* (in preparation) in chapter 13 (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K., in prep.).

5 DISCUSSION

In this study, we systematically investigated voice-identity processing abilities in subjects with unilateral brain lesions. On the group level, the ability to recognize unfamiliar voices was comparable in patients having a lesion in the right hemisphere and patients with lesions in the left hemisphere. Statistical analysis did not reveal significant performance differences between right- and left hemispheric patients. This result contrasted previous traditional lesion studies on voice-identity processing that found voice-identity processing deficits in right brain-damaged patients as compared to left brain-damaged patients (Van Lancker and Canter, 1982; Van Lancker et al., 1989; Neuner and Schweinberger, 2000; Lang et al., 2009).

However, when investigating voice-identity processing abilities on the single subject level, we identified three patients with a selective deficit in voice-identity processing. Intriguingly, lesion locations of these three patients were all restricted to the right hemisphere. More specifically, lesions were assigned to the right temporal lobe including the STG and MTG which are known to be voice-sensitive (Belin et al., 2000, 2002; Kriegstein and Giraud, 2004; Formisano et al., 2008). Lesion location and impairment pattern in these three subjects supports earlier findings in brain damaged patients that described a predominant role of the right hemisphere in human voice-identity processing (Van Lancker and Canter, 1982; Van Lancker et al., 1989; Neuner and Schweinberger, 2000; Lang et al., 2009).

5.1 Voice-Identity Processing Deficits after Brain Damage

Our test battery was designed to investigate voice-identity processing in combination with control tests assessing the specificity of a given voice-identity processing deficit. Although the comparison of voice-identity processing performances between right- and left hemisphere patient group did not reveal significant differences, there were tendencies for a weaker voice processing performance in the right hemispheric patients. We speculate that the following factors could explain comparable group performances: (i) large variance of behavioral data in the voice-identity tests and (ii) the rather small sample size for each hemisphere. Interestingly, in the extended group sample included in the VLSM analysis, the group comparison between patients having lesion in the right

and left hemisphere was significant for the newly-learned voice recognition tests. Further, the large variance in performance within the right-hemispheric sample suggests that rather specific subregions within the right hemisphere are involved in voice-identity processing than the complete right hemisphere. Even though results are limited to the single subject level, they still provide evidence for a selective voice-identity processing deficit which has been shown as a developmental and acquired deficit before (Neuner and Schweinberger, 2000; Roswandowitz et al., 2014). On the other hand, there were patients with combined face- and voice recognition deficits in our sample, a co-occurrence that is in line with earlier clinical findings (Van Lancker and Canter, 1982).

VD patients did not perform significantly weaker on the familiarity decision making or semantic association for the famous voice recognition task than O patients which partly differs from findings in earlier studies. Some studies found dissociations of familiar voice recognition and unfamiliar voice discrimination (Van Lancker and Canter, 1982; Van Lancker et al., 1989; Neuner and Schweinberger, 2000; Lang et al., 2009). Their results implied right-hemispheric dominance in familiar voice recognition impairment, whereas unfamiliar voice discrimination deficits were found in patients with bilateral lesions (Van Lancker et al., 1989). As our test series did not test for unfamiliar voice discrimination but for the recognition of newly- learned voices respective to famous voice recognition, these results are not exactly comparable. One explanation for the better performance on the famous voice recognition task respective to voice-learning might be a different neuronal representation of familiar and unfamiliar/newly-learned voices. It is possible that different neuronal processes accomplish the access to newly learned or familiar voices. Therefore, newly-learned voice-identity processing might be represented in different brain regions than familiar voice-identity processing (Kriegstein and Giraud, 2004).

Next to patients showing a selective deficit in voice-identity processing, there were also eight patients who were indicative of prosopagnosia as revealed by a standard screening tool for face-recognition deficits, i.e. CFMT. Prosopagnosia is a possible phenomenon after brain damage in patients with lesions in the right temporo-occipital region (Damasio et al., 1982; Lang et al., 2006). Lesion location of these eight patients was in the right hemisphere as well. Although voice and face recognition deficits tend to co-

occur, voice-selective recognition can also occur after brain damage (Van Lancker et al., 1989; Neuner and Schweinberger, 2000).

5.2 Selective Voice-Identity Processing Deficits

Patients presenting voice-identity processing deficits (VD patients) were nine patients with right-hemisphere damage after CVI or tumor surgery. Three VD patients presented selective voice processing deficits and it is essential to remark that these three patients had non- overlapping, right temporal lobe damage (see table 8 for details). Face recognition performance and neuropsychological measures were all intact.

While two other VD patients presented very large lesions of parietal lobe and either occipital or frontal lobe, lesions of patients 1, 2 and 7 (see table 9) were smaller or exclusively subcortical. Besides their common lesion side, lesion locations in this group were thus heterogeneous. Patient 2 showed the weakest face-learning performance in combination with a voice processing deficit. His lesion included the fusiform gyrus which is a region being implicated in face-identity processing (McCarthy et al., 1997). Six patients were impaired in both voice- and face recognition suggesting a general deficit in person recognition.

Up to date, clinical studies were not able to provide neuroanatomical correlates to specific voice processing impairments in patients. The study by Hailstone revealed associations between general person recognition impairment and damage in the right anterior temporal pole (Hailstone et al., 2011). The most recent study by Luzzi et al. found correlations of face and voice recognition deficits and right temporal lobe atrophy (Luzzi et al., 2017c). Here, we found 3 right hemisphere damaged patients with specific voice-identity processing impairment with non-overlapping temporal lobe lesions. Patient 8 had a lesion along the right MTG and presented very selective voice processing deficits. Lesions of patient 3 and 9 also mainly involved the temporal lobe and showed the same selective voice processing deficit. The lesion of patient 3 was located along the MTG and STG and extended to the angular gyrus and the mid occipital gyrus. The MTG and STG correspond to current ideas on neuronal correlates of voice processing in the human brain (Belin et al., 2000; Kriegstein and Giraud, 2004). Patient 9 presented a lesion along the cortex of insula and putamen with tangential points to the superior

temporal cortex. These findings are descriptive, therefore we cannot draw general conclusions. The three cases still underpin the great importance that has been ascribed to the right temporal lobe for voice recognition in healthy subjects (Belin et al., 2004; Kriegstein and Giraud, 2004). Beyond the findings in healthy subjects, VLSM analysis for this patient sample found neuroanatomical evidence for voice-selective areas along the superior temporal sulcus.

5.3 Acoustical Features related to Voice-Identity Processing

In the light of acoustical voice features that enable us to identify a person by voice, we gathered an important finding in this study: The ability to process basic acoustical voice features, i.e. pitch and timbre, and the ability to learn and recognize new voices showed a correlation. In line with earlier studies on acoustical voice processing (Singh and Murry, 1978; Gaudrain et al., 2009; Baumann and Belin, 2010), our finding highlights the importance of these two features for voice-identity processing. In our sample, vocal timbre as compared to vocal pitch was the even more relevant voice individuating feature as it showed a strong correlation with the voice-learning task for the entire patient sample (r= -.596 **). Moreover, VD patients performed significantly weaker on the timbre task than O patients (p = .012). It could be suggested that the inability to perceive and analyze this auditory feature might be one reason contributing to weak performance on the voice-learning tasks.

5.4 Subjective Perception of Voice Recognition Deficits after Brain Injury

Two out of three patients with selective voice deficits showed a clinical deficit, i.e. had noted a decline in their personal voice and face recognition abilities after brain injury. This differs from an observation made by Neuner and Schweinberger where patients had not noted any decline of their person recognition abilities (Neuner and Schweinberger, 2000). Considering their good performance on the face-learning tasks, the question whether their impaired voice recognition mechanisms contribute to a subjective impression of more global person recognition impairment can be raised. In a clinical context, future research could embark the ambivalence of subjective person recognition deficits stated by the patients, and the fact that not all patients with objective deficits on person recognition are aware of their impairment. Former studies, (Neuner and

Schweinberger, 2000) as well as our study, obtained mixed results on patients' self-estimation of an impairment. It remains difficult to evaluate the actual degree of every day difficulties based on self-reports in communication through either voice-, face-, or global person recognition deficits. Eight patients reported a decline in their person recognition abilities on the questionnaire without showing objective deficits in the test performance. When attempting to progress in clinical post-brain injury therapy, would patients with a subjective deficit receive treatment if their symptoms were not detectable in objective test measures? Apparently, testing modalities would need improvement towards a more sensitive and clinically applicable method to consider patients' perception of their own impairment. In the long run, a categorization of person recognition impairment after brain injury and their impact amongst all post- brain damage deficits would be desirable. If patients were aware of their deficits and if screening for deficits was sensitive, it might become easier to eventually develop coping strategies to decrease insecurity in personal interaction.

5.5 Implications for current Models of Person Recognition

How do our results assign to the different models of person recognition? The original model by Bruce and Young proposes entirely segregated pathways for each modality up to the level of modality-free PINS (Bruce and Young, 1986b). The voice-adapted model by Belin suggested functionally dissociated pathways for vocal features that follow basic structural encoding of auditory cues. The vocal pathways interact with the corresponding pathways of the face processing system but are generally segregated (Belin et al., 2004). Diverging from the established idea of PINS as the first multimodal level, recent data from neuroimaging studies propose interaction of visual and auditory information at a very early stage of processing (von Kriegstein et al., 2003; Blank et al., 2011; Schall et al., 2013).

Our study design comprised voice and face recognition tests and one subtest on naming, i.e. semantic association. The semantic association task required associating a photograph with a name/face pair without active verbal production as part of the famous voice recognition test. Semantic association has been suggested to be processed in the left hemisphere involving interaction of specifically mid- and anterior superior

temporal sulcus and the fusiform gyrus (Blank et al., 2011). Our behavioral data showed a tendency for weaker performance in semantic association in the left hemisphere group, but we cannot make reliable assumptions about this modality as statistical analysis did not reveal significant group differences.

The question whether PINs exist as supramodal junctions remains unanswered. Regarding behavioral results from the current patient sample, we can only conclude that on a single subject level selective voice processing impairment is more likely to occur in patients with right-hemispheric damage, specifically lesions in the temporal lobe. For these three patients, deficits of voice recognition are dissociated from face recognition and speech impairment (as they were non-aphasic). Our data support implications that voices of different familiarities are processed in distinct brain regions as VD patients did not show difficulties in famous voice recognition. However, according to our data, we cannot draw definite conclusions on vocal and facial pathways and their level of interaction.

5.6 Shortcomings & Open Questions

In summary, we need to mention certain limitations and open questions for future research when regarding the results from the current data. First of all, we did not conduct our behavioral test battery with a healthy control group because we focused on the group comparisons within the patient group. Further extensions of this study could comprise the testing of a respective matched control group that would allow for a definition of more accurate cut-off values for “weak” performers and provide more reliable proportions for the results of both left-and right hemisphere patients’ behavioral performance.

Additionally, increasing the sample size would certainly sharpen the results by reducing the high variance in the data (as can be seen in the larger sample of the VLSM analysis). Instead of including patients with large lesion sites, behavioral analysis would be more comparable if the patients chosen for the sample had more limited or smaller lesions for exclusive behavioral data analysis. However, as patient recruitment for such a comprehensive test battery is already difficult, even stricter inclusion criteria would be another challenge to face and might reduce the sample size.

In total, the complete test battery took five hours of testing for each patient. We divided the sessions into two à 2.5 hours and took short breaks between each test. By randomizing the order of tests in each patient, we intended to rule out the effect of attention loss over time on the test performance. However, the length of testing was intense for patients and required long attention spans. Including a questionnaire on subjective well-being and concentration span after each testing session could be a possibility to take individual loss of attention into consideration. The data could be considered as a covariate in data analysis.

Neuronal correlates of specific voice processing deficits could only be described in single subjects in the chosen method of data analysis. Imaging analysis as in VLSM as a further step of analysis allows for definite statements on correlations between behavioral deficits and neuronal correlates. This more elaborate analysis showed associations of right temporal lobe damage and voice-identity processing and confirms our findings from a single subject level (Roswandowitz, C., Kappes, C., Obrig, H., von Kriegstein K., in prep.).

5.7 Conclusion

This study characterized voice recognition patterns in patients with unilateral brain lesions. We identified a group of patients evidencing deficits in voice-identity processing. Lesions were all located in the right hemisphere. While most patients had difficulties in voice and face-identity processing, three subjects presented a selective deficit in voice recognition. These three subjects had non-overlapping right temporal lobe lesions mainly locating in the STG and MTG. This finding supports earlier assumptions on neuronal correlates in healthy subjects (Belin et al., 2000; Kriegstein and Giraud, 2004). Further analysis through VLSM confirmed the important role of the right temporal lobe in human voice processing and implies an involvement of the inferior parietal lobe in multimodal person recognition, i.e. face-voice integration.

Compared to earlier patient studies in this field, our study provided an exceptionally detailed and broad test series. Each patient was tested in two testing sessions and completed a complex test battery with multiple subtests on voice-identity processing of different familiarities and control tests (face tests, neuropsychology, audiometry). High resolution brain images were available and allowed for the detection of precise lesion location in patients with selective voice-identity processing deficits. In a clinical context, findings from the current study might support the need for a better understanding of post-stroke non-verbal communication disabilities and their eventual treatment concepts.

6 ZUSAMMENFASSUNG DER ARBEIT

Dissertation zur Erlangung des akademischen Grades *Dr. med.*

Voice-Identity Processing in Patients with Brain Lesions

eingereicht von
Claudia Dominica Kappes, geb. am 29.06.1988 in Münster

angefertigt am Max-Planck-Institut für Kognitions-und Neurowissenschaften Leipzig
und der Tagesklinik für kognitive Neurologie des Universitätsklinikums Leipzig

betreut von
Prof. Dr. med. Hellmuth Obrig und Prof. Dr. med. Katharina von Kriegstein
März 2017

Hintergrund:

Die menschliche Stimme wird in der Fachliteratur als “auditorisches Gesicht” bezeichnet (Belin et al., 2004), weil sie neben der Sprache auch Informationen zu Identität und Emotionen des Sprechers vermittelt, die wir in der alltäglichen Kommunikation mühelos wahrnehmen und verarbeiten. Zerebrale Pathologien, beispielsweise ischämische Hirninfarkte oder Hämorrhagien, können in der Folge verschiedene Kommunikationsdefizite verursachen. Ein bedeutsames Kommunikationsdefizit auf sprachlicher Ebene ist die *Aphasie*. Defizite der Stimmerkennung als eine Entität der nicht-sprachlichen Ebene werden durch den Begriff *Phonagnosie* definiert. *Phonagnosie* beschreibt ein Defizit der Stimmidentifizierung einschließlich der Analyse akustischer vokaler Signale, dem Wiedererkennen bekannter Stimmen und der semantischen Assoziation einer erkannten Stimme (Roswandowitz C, Maguinness C, von Kriegstein, in rev.). Klinische Studien wiesen die Existenz von Stimmerkennungsdefiziten als eine mögliche Folge zerebraler Läsionen nach (Van Lancker and Canter, 1982; Van Lancker et al., 1989; Neuner and Schweinberger, 2000; Lang et al., 2009; Hailstone et al., 2011). Hierbei wurden insbesondere Läsionen der rechten Hemisphäre als zugrundeliegende

neuronale Repräsentationen hervorgehoben, allerdings gelang bisher keine exakte Lokalisierung der betroffenen Hirnregionen bei isolierten Stimmerkennungsdefiziten. In funktionellen MRT-Studien an gesunden Probanden zeigten sich stimmspezifische Areale entlang des rechten superioren temporalen Gyrus und Sulcus (STG/S) (Belin et al., 2000; von Kriegstein et al., 2003; Kriegstein and Giraud, 2004).

Zielsetzung:

Ziel der vorliegenden Patientenstudie war es, mögliche isolierte Stimmerkennungsdefizite als Folge einer zerebralen Läsion nachzuweisen und zu charakterisieren. In einem zweiten Schritt widmete sich die Studie der Frage nach den neuronalen Korrelaten von Stimmerkennungsdefiziten. Wir stellten die Hypothesen auf, dass Stimmerkennungsdefizite (i) häufiger bei Patienten mit rechtshemisphärischen Läsionen und (ii) darüber hinaus als isoliertes Stimmerkennungsdefizit gegenüber kombinierten Defiziten von Stimm- und Gesichtserkennung auftreten können. Die Untersuchung von neuronalen Korrelaten dieser Defizite wurde in einer weiterführenden Analyse mittels Voxel-based lesion symptom mapping (VLSM) vorgenommen (Roswandowitz, C., Kappes, C., Obrig, H., von Kriegstein K., in prep.).

Material und Methoden:

40 Patienten der Tagesklinik für kognitive Neurologie der Universität Leipzig nahmen an der Studie teil. Alle Patienten wiesen unilaterale Hirnläsionen (n = 14 links, 24 rechts) auf, die entweder Folge eines cerebrovaskulären Ereignisses oder einer Tumorextraktion waren. Wir führten eine umfangreiche experimentelle Testreihe durch, die insbesondere der Stimmerkennung (Stimmlerntests und Tests zur Erkennung bekannter Stimmen) galt. Außerdem wurde die Kontrollmodalität der Gesichtererkennung und die Verarbeitung akustischer vokaler Signale (Pitch und Timbre) überprüft. Die individuelle Patientenwahrnehmung zur Stimm- und Gesichtererkennung erhoben wir in einem Fragebogen. Wir analysierten die Daten in IBM SPSS 22, für die Gruppenvergleiche wendeten wir sowohl parametrische als auch nicht-parametrische Tests, Varianzanalysen und bivariate Korrelationen an. In einem weiterführenden Teil der Studie wurden die behavioralen Daten und strukturelle MRTs anhand von Voxel-based lesion symptom mapping (VLSM) analysiert.

Ergebnisse:

In der Datenanalyse fanden sich im Gruppenvergleich der Patienten mit rechts- bzw. linkshemisphärischen Läsionen keine signifikanten Unterschiede in den Tests zur Stimmerkennung. Allerdings wiesen 9 Patienten, deren Läsionen ausschließlich rechtshemisphärisch lokalisiert waren, Stimmererkennungsdefizite auf. Die Lokalisation der Läsionen innerhalb der rechten Heimsphäre war heterogen.

Während sechs Patienten dieser Gruppe ein kombiniertes Defizit der Gesichter- und Stimmerkennung zeigten, fand sich bei drei Patienten ein isoliertes Defizit der Stimmerkennung. Wir charakterisieren in der vorliegenden Arbeit das spezifische Verhaltensmuster und die Lokalisation der Läsionen dieser drei Patienten, die alle eine Beteiligung des rechten Temporallappens aufwiesen. Im Hinblick auf grundlegende Mechanismen der Stimmverarbeitung konnte insbesondere Timbre als relevantes akustisches Stimmsignal zur Erkennung neu erlernter Stimmen identifiziert werden. In der weiterführenden Analyse mittels VLSM wurden Assoziationen von (i) selektiven Defiziten der Stimmerkennung mit Läsionen im rechten Temporallappen sowie (ii) der Stimm-Gesichter-Integration im rechten inferioren Parietallappen nachgewiesen.

Schlussfolgerungen:

Die vorliegende Studie hebt auf der Grundlage des untersuchten Patientenkollektivs die bedeutsame Rolle der rechten Hemisphäre bei der Stimmerkennung hervor. Wir identifizierten drei Patienten mit isolierten Stimmerkennungsdefiziten, deren Läsionen sich im rechten Temporallappen befanden. Dieses Ergebnis stützt bisherige Evidenz zur Stimmverarbeitung an gesunden Probanden (Belin et al., 2000; Kriegstein and Giraud, 2004). Die weiterführende VLSM-Analyse, auf der Grundlage des vorliegenden Patientenkollektivs, charakterisiert spezifische Areale des rechten Temporallappens und inferioren Parietallappens als neuronale Korrelate defizitärer Stimmerkennung. In Erweiterung bisheriger klinischer Evidenz liefert die vorliegende Studie neue Erkenntnisse zu neuronalen Korrelaten von isolierten Stimmerkennungsdefiziten und Defiziten der Stimm-Gesichter-Integration (Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K., in prep.). Im klinischen Kontext erlaubt die Studie einen weiteren Schritt zum besseren Verständnis von nonverbalen Kommunikationsdefiziten, insbesondere Stimmerkennungsschwierigkeiten, nach zerebralen Läsionen.

7 REFERENCES

A Kawahara H, A Masuda-Katsuse I, A de Cheveigné A (1999) Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. J Speech Commun.

Alba-Ferrara L, Weis S, Damjanovic L, Rowett M, Hausmann M (2012) Voice identity recognition failure in patients with schizophrenia. J Nerv Ment Dis 200:784–790.

Assal G, Zander E, Kremin H, Buttet J (1976) [Voice discrimination in patients with cerebral cortical lesions]. Schweiz Arch Neurol Neurochir Psychiatr 119:307–315.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003a) Voxel-based lesion-symptom mapping. Nat Neurosci 6:448–450.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003b) Voxel-based lesion-symptom mapping. Nat Neurosci 6:448–450.

Baumann O, Belin P (2010) Perceptual scaling of voice identity: common dimensions for different vowels and speakers. Psychol Res 74:110–120.

Belin P, Fecteau S, Bedard C (2004) Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105–2109.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17–26.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Bestelmeyer PE, Belin P, Grosbras MH (2011) Right temporal TMS impairs voice detection. Curr Biol 21:R838-9.

Blank H, Anwander A, von Kriegstein K (2011) Direct structural connections between voice- and face-recognition areas. J Neurosci 31:12906–12915.

Boersma P, Weenink D (2001) Praat, a system for doing phonetics by computer.

Borovsky A, Saygin AP, Bates E, Dronkers N (2007) Lesion correlates of conversational speech production deficits. Neuropsychologia 45:2525–2533.

Brownsett SL, Warren JE, Geranmayeh F, Woodhead Z, Leech R, Wise RJ (2014) Cognitive control and its impact on recovery from aphasic stroke. Brain 137:242–254.

Bruce V, Young A (1986a) Understanding face recognition. Br J Psychol 77:305–327.

Bruce V, Young A (1986b) Understanding face recognition. Br J Psychol 77 (Pt 3):305–327.

Capilla A, Belin P, Gross J (2013) The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cereb Cortex 23:1388–1395.

Charest I, Pernet CR, Rousselet GA, Quinones I, Latinus M, Fillion-Bilodeau S, Chartrand JP, Belin P (2009) Electrophysiological evidence for an early processing of human

voices. BMC Neurosci 10:127.

Chhabra S, Badcock JC, Maybery MT, Leung D (2012) Voice identity discrimination in schizophrenia. Neuropsychologia 50:2730–2735.

Collins SA, Missing C (2003) Vocal and visual attractiveness are related in women. Anim Behav 65:997–1004.

Damasio AR, Damasio H, Van Hoesen GW (1982) Prosopagnosia: anatomic basis and behavioral mechanisms. Neurology 32:331–341.

DeCasper AJ, Fifer WP (1980) Of human bonding: newborns prefer their mothers' voices. Science (80-) 208:1174–1176.

Dronkers NF, Wilkins DP, Van Valin RD, Redfern BB, Jaeger JJ (2004) Lesion analysis of the brain areas involved in language comprehension. Cognition 92:145–177.

Duchaine B, Nakayama K (2006) The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 44:576–585.

Ellis HD, Jones DM, Mosdell N (1997) Intra- and inter-modal repetition priming of familiar faces and voices. Br J Psychol 88:143–156.

Fant G (1960) Acoustic theory of speech production. The Hague, Netherlands.

Fimm B, A Zimmermann P (2001) Testbatterie zur Aufmerksamkeitsprüfung (TAP)-Version 1.6. Herzogenrath: Psytest.

Fitch WT (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. J Acoust Soc Am 102:1213–1222.

Fleming D, Giordano BL, Caldara R, Belin P (2014) A language-familiarity effect for speaker discrimination without comprehension. Proc Natl Acad Sci 111:13795–13798.

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science (80-) 322:970–973.

Gainotti G (2007) Face familiarity feelings, the right temporal lobe and the possible underlying neural mechanisms. Brain Res Rev 56:214–235.

Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, Schweinberger SR, Warren JD, Duchaine B (2009) Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 47:123–131.

Gaudrain E, Li S, Ban VS, Patterson RD (2009) The role of glottal pulse rate and vocal tract length in the perception of speaker identity. In: INTERSPEECH, pp 148–151.

Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthelemy C, Brunelle F, Samson Y, Zilbovicius M (2004) Abnormal cortical voice processing in autism. Nat Neurosci 7:801–802.

Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD (2010) Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia 48:1104–1114.

Hailstone JC, Ridgway GR, Bartlett JW, Goll JC, Buckley AH, Crutch SJ, Warren JD (2011) Voice processing in dementia: a neuropsychological and neuroanatomical analysis.

Brain 134:2535–2547.

Härting C, Markowitsch H, Neufeld H (2000) WMS-R. … Dtsch Adapt der ….

Henseler I, Regenbrecht F, Obrig H (2014) Lesion correlates of patholinguistic profiles in chronic aphasia: comparisons of syndrome-, modality- and symptom-level assessment. Brain 137:918–930.

Ives DT, Smith DR, Patterson RD (2005) Discrimination of speaker size from syllable phrases. J Acoust Soc Am 118:3816–3822.

Kaernbach C (1991) Simple adaptive testing with the weighted up-down method. Percept Psychophys 49:227–229.

Kawahara H, Morise M, Takahashi T, Nisimura R, Irino T, Banno H (2008) Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp 3933–3936. IEEE.

Kertesz A, McCabe P (1977) Recovery patterns and prognosis in aphasia. Brain 100 Pt 1:1–18.

Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z (2003) Effects of experience on fetal voice recognition. Psychol Sci 14:220–224.

Kreiman J, Sidtis DVL (2011) Foundations of voice studies: An interdisciplinary approach to voice production and perception. Wiley. com.

Kriegstein K V, Giraud AL (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG (2009) Voice recognition in aphasic and non-aphasic stroke patients. J Neurol 256:1303–1306.

Lang N, Baudewig J, Kallenberg K, Antal A, Happe S, Dechent P, Paulus W (2006) Transient prosopagnosia after ischemic stroke. Neurology 66:916.

Lass NJ, Hughes KR, Bowyer MD, Waters LT, Bourne VT (1976) Speaker sex identification from voiced, whispered, and filtered isolated vowels. J Acoust Soc Am 59:675–678.

Laver J (1980) The phonetic description of voice quality. Cambridge Stud Linguist London 31:1–186.

Lavner Y, Gath I, Rosenhouse J (n.d.) The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels.

Levy DA, Granot R, Bentin S (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40:291–305.

Luzzi S, Baldinelli S, Ranaldi V, Fabi K, Cafazzo V, Fringuelli F, Silvestrini M, Provinciali L, Reverberi C, Gainotti G (2017a) Famous faces and voices: Differential profiles in early right and left semantic dementia and in Alzheimer's disease. Neuropsychologia 94:118–128.

Luzzi S, Baldinelli S, Ranaldi V, Fabi K, Cafazzo V, Fringuelli F, Silvestrini M, Provinciali L, Reverberi C, Gainotti G (2017b) Famous faces and voices: Differential profiles in early right and left semantic dementia and in Alzheimer's disease.

Neuropsychologia 94:118–128.

Luzzi S, Baldinelli S, Ranaldi V, Fabi K, Cafazzo V, Fringuelli F, Silvestrini M, Provinciali L, Reverberi C, Gainotti G (2017c) Famous faces and voices: Differential profiles in early right and left semantic dementia and in Alzheimer's disease. Neuropsychologia 94:118–128.

Macmillan NA, Creelman CD (2004) Detection Theory: A User's Guide.

Mathias SR, von Kriegstein K (2014) How do we recognise who is speaking? Front Biosci (Schol Ed) 6:92–109.

McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. J Cogn Neurosci 9:605–610.

Neuner F, Schweinberger SR (2000) Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn 44:342–366.

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

Pascolini D, Smith A (2009) Hearing Impairment in 2008: a compilation of available epidemiological studies. Int J Audiol 48:473–485.

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

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

Roswandowitz, C., Kappes, C., Obrig, H. von Kriegstein K (2017) Voice-identity recognition deficits are induced by lesions in the temporal and inferior parietal lobe. Prep.

Roswandowitz C, Mathias SR, Hintz F, Kreitewolf J, Schelinski S (2014) Report Two Cases of Selective Developmental Voice-Recognition Impairments. Curr Biol:1–6.

Roswandowitz C, Maguinness C von KK (n.d.) Deficits in voice-identity processing: acquired and developmental phonagnosia. In: The Oxford Handbook of Voice Perception, under review.

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

Schall S, Kiebel SJ, Maess B, von Kriegstein K (2015) Voice Identity Recognition: Functional Division of the Right STS and Its Behavioral Relevance. J Cogn Neurosci 27:280–291.

Schelinski S, Borowiak K, von Kriegstein K (2016) Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. Soc Cogn Affect Neurosci 11:1812–1822.

Schelinski S, Borowiak K, von Kriegstein K (n.d.) Neural processing of voices in autism spectrum disorder. In: 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM).

Schmidt KH, A Metzler P (1992) WST-Wortschatztest. Handbuch, Göttingen Beltz Test.

Schuri U, Benz R (2000) Gesichter-Namen-Lerntest:(GNL).

Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001) The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 124:804–815.

Shilowich BE, Biederman I (2016) An estimate of the prevalence of developmental phonagnosia. Brain Lang 159:84–91.

Singh S, Murry T (1978) Multidimensional classification of normal voice qualities. J Acoust Soc Am 64:81–87.

Skuk VG, Schweinberger SR (2013) Influences of Fundamental Frequency, Formant Frequencies, Aperiodicity and Spectrum Level on the Perception of Voice Gender. J Speech Lang Hear Res.

Smith DR, Patterson RD (2005) The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. J Acoust Soc Am 118:3177–3186.

Smith DRR, Patterson RD, Turner R, Kawahara H, Irino T (2005) The processing and perception of size information in speech sounds. J Acoust Soc Am 117:305–318.

Van Lancker D, Kreiman J (1987) Voice discrimination and recognition are separate abilities. Neuropsychologia 25:829–834.

Van Lancker DR, Canter GJ (1982) Impairment of voice and face recognition in patients with hemispheric damage. Brain Cogn 1:185–195.

Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH (1988) Phonagnosia: A Dissociation Between Familiar and Unfamiliar Voices. Cortex 24:195–209.

Van Lancker DR, Kreiman J, Cummings J (1989) Voice perception deficits: neuroanatomical correlates of phonagnosia. J Clin Exp Neuropsychol 11:665–674.

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

von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD (2010) How the human brain recognizes speech in the context of changing speakers. J Neurosci 30:629–638.

von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD (2006) Processing the acoustic effect of size in speech sounds. Neuroimage 32:368–375.

von Kriegstein K. Roswandowitz C MC, C M, K. von K (2017) Deficits in voice-identity processing: acquired and developmental phonagnosia. Oxford Handb Voice Perception, under Rev.

Xu X, Biederman I, Shilowich BE, Herald SB, Amir O, Allen NE (2015) Developmental phonagnosia: Neural correlates and a behavioral marker. Brain Lang 149:106–117.

Xu Y, Lee A, Wu W-L, Liu X, Birkholz P (2013) Human vocal attractiveness as signaled by body size projection. PLoS One 8:e62397.

8 SUPPLEMENTARY MATERIAL

8.1 Individual Demographical Data

Subject	Age	Sex	Education years	Etiology	Hemisphere	Months since Onset
2008	50	f	12	ICH	r	-
2034	62	m	12	ICH	l	57
2047	22	f	10	ICH	r	54
2075	52	m	10	isch	l	47
2086	50	m	12	ICH	r	52
2096	43	m	10	isch	r	48
2253	58	m	12	isch	r	42
2337	45	f	10	isch	r	39
2390	57	m	10	SAB	r	excluded
2406	46	f	10	SAB	r	40
2456	56	m	10	isch	r	28
2458	61	f	10	isch	r	33
2470	54	f	10	isch	l	30
2475	31	f	10	tumor	l	31
2486	63	m	12	isch	r	31
2526	35	m	10	isch	r	31
2550	56	f	10	ICH	r	24
2552	53	f	11	isch	l	25
2583	49	m	10	isch	r	12
2637	50	f	10	isch	r	35
2678	29	m	12	SAB	l	24
2710	23	m	10	isch	r	14
2712	57	f	10	isch	l	18
2717	51	f	10	ICH	r	20
2764	31	f	10	SAB	r	17
2770	57	f	10	isch.	l	11
2779	47	f	10	SAB	r	21
2800	44	f	10	isch	l	2
2802	49	f	12	isch	l	10
2815	58	m	12	isch	l	9
2828	29	m	10	isch	l	3
2847	40	f	10	isch	l	excluded
2849	55	m	10	Tumor	l	5
2851	50	m	8	SAB	r	14
2857	48	m	10	isch	l	11
2870	46	m		isch	l	excluded
2899	32	f	12	Tumor	r	12
2905	47	m	10	isch	r	7
2924	58	f	12	isch	r	10
2989	57	f	10	isch	r	13

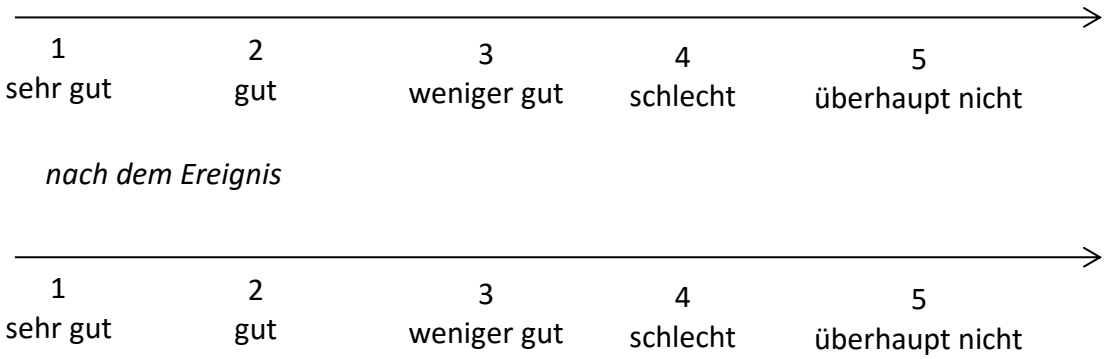
Table 10: Individual demographical data. *Isch= ischemic, ICH = intracerebral hemorrhage, SAB = subarachnoid hemorrhage

8.2 Questionnaire on Person Recognition

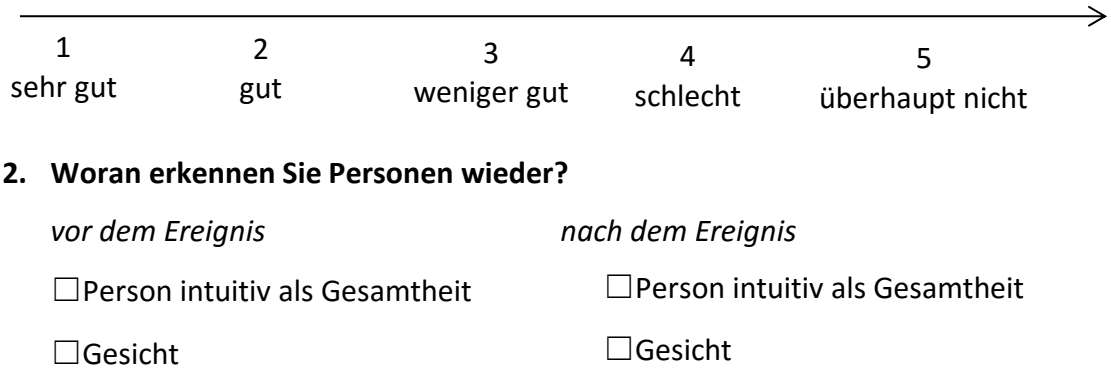
Fragen zur Personenerkennung

1. **Wie gut erkennen Sie bekannte Personen wieder?**

   vor dem Ereignis



   nach dem Ereignis



2. **Woran erkennen Sie Personen wieder?**

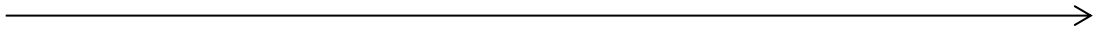
vor dem Ereignis	*nach dem Ereignis*
☐Person intuitiv als Gesamtheit	☐Person intuitiv als Gesamtheit
☐Gesicht	☐Gesicht
☐Stimme	☐Stimme
☐Kleidung, Schmuck, Frisur, Brille	☐Kleidung, Schmuck, Frisur, Brille
☐Körperhaltung, Gang	☐Körperhaltung, Gang
☐besondere Merkmale: Narbe, Hinken u	☐ besondere Merkmale: Narbe. Hinken
☐andere Merkmale:	

3. Wenn Sie nur die Stimme einer Ihnen bekannten Person hören, wie gut würden Sie diese Person erkennen?

vor dem Ereignis

1	2	3	4	5
sehr gut	gut	weniger gut	schlecht	überhaupt nicht

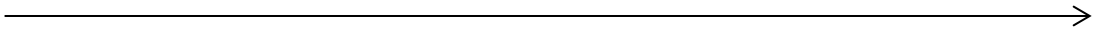
nach dem Ereignis



1	2	3	4	5
sehr gut	gut	weniger gut	schlecht	überhaupt nicht

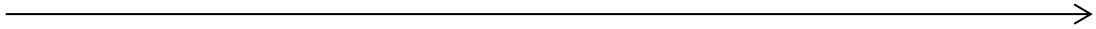
4. Wenn Sie nur das Gesicht einer Ihnen bekannten Person sehen, wie gut würden Sie diese Person erkennen?

vor dem Ereignis



1	2	3	4	5
sehr gut	gut	weniger gut	schlecht	überhaupt nicht

nach dem Ereignis



1	2	3	4	5
sehr gut	gut	weniger gut	schlecht	überhaupt nicht

8.3 Behavioral Test Results

NPS (%) (1 SD)	Voice Learning (%)	Face Learning (%)	Famous Voices d`	Famous Voices c	Famous Voices Naming (%)	Pitch (cent)	Timbre (SER)
47.95 *(14.02)*	59.71 *(9.88)*	61.65 *(14.99)*	1.52 *(1.30)*	-2.7 *(.59)*	70.92 *(17.49)*	124.14 *(63.40)*	9.39 *(4.48)*

Table 11: Performance for the entire patient sample on all behavioral tests. NPS= Neuropsychology.

8.4 Test Performance of VD and O Patients: Basic Acoustic Voice Properties

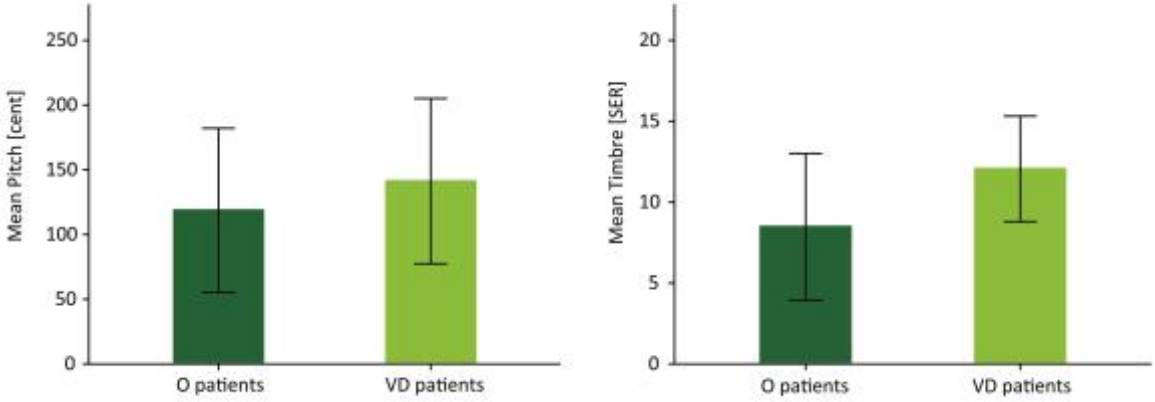


Figure 14: Performance on auditory control tests for VD and O patients. Left: Mean Pitch (cent), Right: Mean Timbre (SER). Error bars represent +/- 1 SD.

8.5 Test Performance of VD and O Patients: Control Modality Face Processing

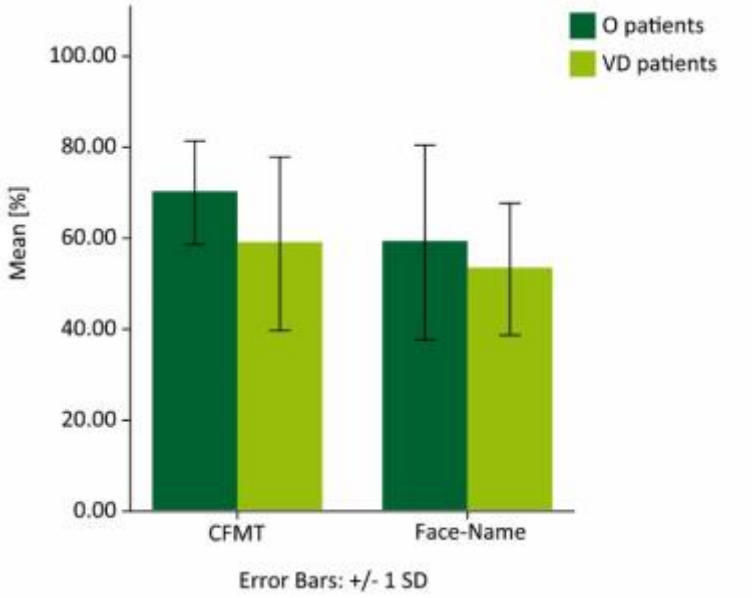


Figure 15: Performance of VD and O patients on the face control tests: Cambridge Face Memory Test (CFMT) and Face-Name test (%). Error bars represent +/ - 1 SD.

8.6 Correlations

		NPS (PR)	Overall voice-learning	d'	c value	Naming	Face-learning	Pitch	Timbre
NPS (PR)	Pearson Correlation	1	0.113	-0.243	0.04	-0.14	0.049	-0.418	-0.175
	Sig. (2-tailed)		0.525	0.166	0.824	0.429	0.784	0.014	0.329
	N	34	34	34	34	34	34	34	33
Overall voice-learning	Pearson Correlation	-0.001	1	-0.069	-0.134	0.232	0.321	-0.365*	-0.545**
	Sig. (2-tailed)	0.997		0.686	0.431	0.167	0.053	0.026	0.001
	N	37	37	37	37	37	37	37	36
d'	Pearson Correlation	-.103	-0.102	1	0.396*	0.511**	0.144	0.147	-0.07
	Sig. (2-tailed)	.549	0.553		0.017	0.001	0.403	0.392	0.69
	N	36	36	36	36	36	36	36	35
c value	Pearson Correlation	-.041	-0.123	0.396*	1	-0.404 *	-0.096	0.033	-0.012
	Sig. (2-tailed)	.811	0.474	0.017		0.014	0.578	0.848	0.946
	N	36	36	36	36	36	36	36	35
Naming	Pearson Correlation	-.054	0.175	0.511**	-0.404**	1	0.324	0.151	-0.135
	Sig. (2-tailed)	.755	0.308	0.001	0.014		0.054	0.378	0.439
	N	36	36	36	36	36	36	36	35
Face-Learning	Pearson Correlation	0.15	0.321	0.096	-0.077	0.141	1	0.103	-0.061
	Sig. (2-tailed)	0.376	0.053	0.57	0.652	0.406		0.545	0.722
	N	37	37	37	37	37	37	37	36
Pitch	Pearson Correlation	-0.177	-.365 *	.111	.045	.030	.103	1	.494**
	Sig. (2-tailed)	0.295	.026	.515	.790	.858	.545		.002
	N	37	37	37	37	37	37	37	36
Timbre	Pearson Correlation	-0.046	-0.596**	0.059	0.031	-0.16	-0.043	0.591**	1
	Sig. (2-tailed)	0.792	0	0.736	0.861	0.358	0.805	0	
	N	35	35	35	35	35	35	35	35

Table 12: Correlations of all tests for the entire patient sample.

9 LIST OF FIGURES

10 LIST OF TABLES

11 ERKLÄRUNG ÜBER DIE EIGENSTÄNDIGE ABFASSUNG DER ARBEIT

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar eine Vergütung oder geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt an der Entstehung der vorliegenden Arbeit beteiligt waren. Die aktuellen gesetzlichen Vorgaben in Bezug auf die Zulassung der klinischen Studien, die Bestimmungen des Tierschutzgesetzes, die Bestimmungen des Gentechnikgesetzes und die allgemeinen Datenschutzbestimmungen wurden eingehalten. Ich versichere, dass ich die Regelungen der Satzung der Universität Leipzig zur Sicherung guter wissenschaftlicher Praxis kenne und eingehalten habe.

Datum Unterschrift

12 LEBENSLAUF

Claudia Kappes
Könneritzstraße 25
04229 Leipzig

0176-85001586
claudia.d.kappes@gmx.de

Geburtsdatum	29.06.1988
Geburtsort	Münster
Familienstand	Ledig
Nationalität	Deutsch

AUSBILDUNG

1994-2007	SCHULISCHE AUSBILDUNG Gymnasium Augustinianum Greven, Abitur Sommer 2007
2008-2015	STUDIUM AN DER UNIVERSITÄT LEIPZIG Humanmedizin, Physikum im September 2010 2. Staatsexamen im Dezember 2015
Seit 01/2016	ÄRZTIN IN WEITERBILDUNG Klinik und Poliklinik für Kinder- und Jugendmedizin, Universitätsklinikum Leipzig

KLINISCHE AUSBILDUNG

11/2014- 10/2015	Praktisches Jahr
1. Tertial	Viszeralchirurgie, Ev.Diakonissenkrankenhaus Leipzig Kinderchirurgie, Schneider Children's Medical Center of Israel, Tel Aviv, Israel
2. Tertial	Pädiatrie, Universitätsklinik für Kinder-und Jugendmedizin, Leipzig Allgemeinpädiatrie Pädiatrische Onkologie
3. Tertial	Innere Medizin, St. Elisabeth Krankenhaus Leipzig

03/2013	Famulatur Psychiatrie, Universitäre Psychiatrische Dienste (UPD) Bern, Schweiz
09/2012	Famulatur Innere Medizin, MJH Greven
05-06/2012	Famulatur pädiatrische Notaufnahme, Hôpital Mère-Enfant, Lyon, Frankreich
03-04/2012	Famulatur Neurologie, CHU Lyon, Frankreich
08-09/2011	Famulatur Pädiatrie, Emory University Hospitals, Atlanta, GA, USA

WISSENSCHAFTLICHE ARBEIT

seit 02/2013	*Promotion* am Max-Planck Institut für Kognitions- und Neurowissenschaften und der Tagesklinik für kognitive Neurologie der Universität Leipzig Thema: *Voice-Identity Processing in Patients with Brain Lesions* Arbeitsgruppe *Neural mechanisms of human communication*, Prof. Dr. Katharina von Kriegstein

AUSLANDSAUFENTHALTE

01/2015 – 03/2015	ISRAEL PJ Kinderchirurgie am *Schneider children's medical center of Israe*l, Tel Aviv University
01/2012 - 08/2012	FRANKREICH Auslandssemester an der Université Claude-Bernard, Lyon
08-09/2011	USA Famulatur Pädiatrie an den Emory University Hospitals, Atlanta
09/2007 - 09/2008	Freiwilliges Soziales Jahr in der Jüdischen Gemeinde, Prag

WEITERBILDUNG

Sonographie	Grundkurs der Sonographie an der Universität Leipzig, Leitung Prof. Keim, WS 2013/14
Statistik	SPSS Kurs des IMISE an der Universität Leipzig, SS 2014

WEITERE TÄTIGKEITEN

2013-2014	Studentische Hilfskraft am Max-Planck-Institut, Studie „Hippocampal Plasticity“
2013-2014	Studentische Hilfskraft am Carl-Ludwig-Institut für Physiologie, Universität Leipzig
02-04/ 2013	Studentische Hilfskraft auf der Intensivstation der Universitätsklinik für Kinder und Jugendmedizin, Leipzig
2011-2012	Studentische Hilfskraft im Schlaflabor der Pneumologie, Universitätsklinikum Leipzig
2011-12	Mitglied im Vorstand des Leipziger Universitätsorchesters, Ressort “Presse- und Öffentlichkeitsarbeit”
Seit 10/2008	Cellistin im Leipziger Universitätsorchester
09/2007-09/2008	Freiwilliges Soziales Jahr: Aktion Sühnezeichen Friedensdienste Offene Altenarbeit mit Holocaustüberlebenden und Mitarbeit im jüdischen Kindergarten der Jüdischen Gemeinde, Prag

SPRACHKENNTNISSE

Englisch: fließend in Wort und Schrift
Französisch: fließend in Wort und Schrift
Tschechisch: gute Kenntnisse

PERSÖNLICHE INTERESSEN

Musik: Cello und Klavier
Literatur und Theater
Sprachen und Reisen

13 MANUSKRIPT DER PUBLIKATION

13.1 Voice-identity recognition deficits are induced by lesions in the temporal and inferior parietal lobe

Author names	Claudia Roswandowitz [a,b] (roswandowitz@cbs.mpg.de) Claudia Kappes [a] Hellmuth Obrig [a, c] Katharina von Kriegstein [a, d]
Author affiliations	[a] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany [b] International Max Planck Research School on Neuroscience of Communication, Stephanstraße 1a, 04103 Leipzig, Germany [c] Clinic for Cognitive Neurology, University Hospital Leipzig, Germany [d] Humboldt University zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany
Corresponding author	Claudia Roswandowitz Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany Email: roswandowitz@cbs.mpg.de Fax: +49 (0) 341-9940-2606
Any conflict of interest	The authors declare no competing financial interests.

Number of Figures	5
Number of Tables	4 (max 8 including tables and figures, BRAIN)
Supplementary Material	Supplementary Methods, Results, References
Number of words	*Abstract:* 244 *Main text:* 7100

Abstract

Voice-identity recognition is an important skill for social interactions. It is by-and-large unknown how this skill is accomplished by the human brain. In the past decade a wealth of neuroimaging studies have shown that a central structure involved in voice-identity recognition is the right temporal lobe. However, neuropsychological case studies reported that voice-identity recognition deficits in patients with brain lesions are associated with lesions in the right inferior parietal lobe. The aim of the present study was to work towards resolving the discrepancy between neuroimaging studies and patient lesion case reports to get a better understanding, which structures in the human brain are critical for voice-identity recognition. To do this we performed a voxel based lesion symptom mapping (VLSM) study on 58 patients with unilateral focal brain lesions. The study included a comprehensive behavioural test battery, neuropsychological assessment and high-resolution structural brain images. The VLSM analysis revealed three key findings. (i) We identified a strong association between lesions in the temporal and right inferior parietal lobe and voice-identity recognition deficits. (ii) Of these two structures, only the right temporal lobe remained significant when we controlled for face-recognition performance indicating a high voice sensitivity of the right temporal lobe. (iii) The right inferior parietal lobe was particularly involved in tasks which required integration of voice and face information. The results imply that the right temporal lobe is an obligatory structure for voice-identity recognition, while the inferior parietal lobe is a facultative component of voice-identity recognition.

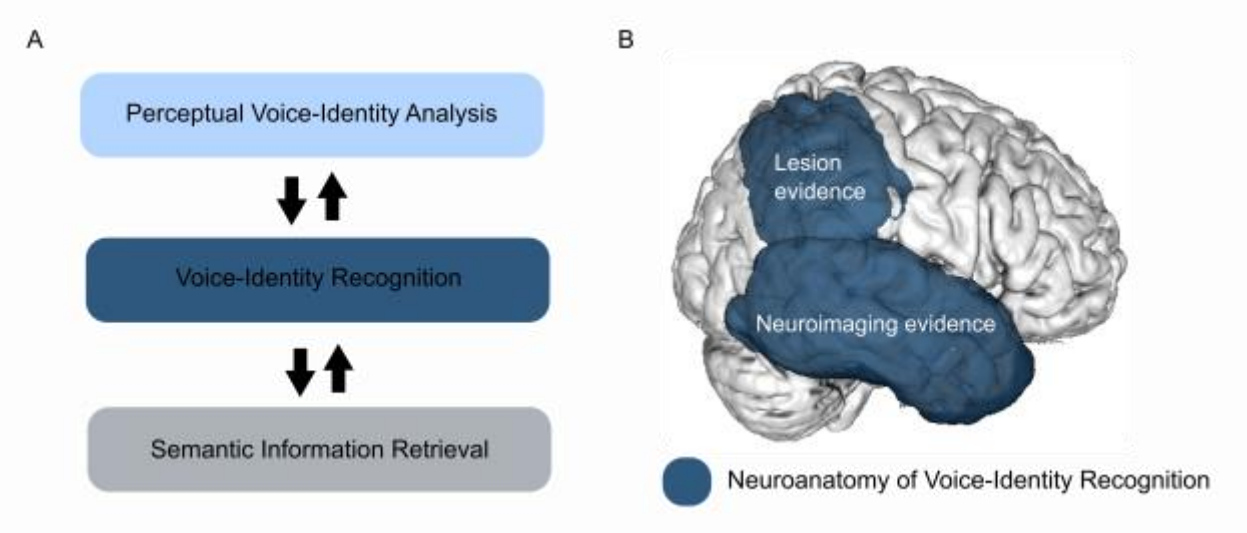
Keywords: voice recognition, voxel-based lesion symptom mapping, temporal lobe, parietal lobe

Introduction

The ability to recognise the identity of other people is a key skill for successful human communication. Impairments in person 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 recognising a person by voice can be acquired due to brain lesions or neurodegenerative diseases (e.g., Van Lancker and Canter, 1982; Neuner and Schweinberger, 2000; Hailstone *et al.*, 2011). They can also be symptoms of developmental and psychiatric disorders such as autism spectrum disorder and schizophrenia (e.g. Boucher *et al.*, 1998; Garrido *et al.*, 2009; Alba-Ferrara *et al.*, 2012; Mou *et al.*, 2013; Roswandowitz *et al.*, 2014; Schelinski *et al.*, 2016a; Schelinski *et al.*, 2016b).

A deficit in voice-identity processing has been first described in patients with brain lesions and has been termed phonagnosia (Van Lancker and Canter, 1982). The clinical case studies have shown that brain lesions can lead to relatively selective impairments in recognising familiar person by voice, i.e. voice-identity recognition. In these cases voice-identity recognition deficits were dissociated from intact face-identity recognition skills (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Neuner and Schweinberger, 2000), from relatively intact language skills (Assal *et al.*, 1976; Assal *et al.*, 1981; Lang *et al.*, 2009), and from intact perceptual voice analysis stages, i.e. discriminating whether voices are same or different (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). A cognitive model of voice-identity processing is shown in Fig. 1A. These findings suggested a dedicated neural substrate for voice-identity recognition that is not involved to the same extent in other person recognition, language perception skills or even the perceptual analysis of voice identity features before identity recognition takes place. A few studies report detailed lesion locations for acquired phonagnosia cases (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). Based on these studies, a prime candidate area for voice-identity recognition is the right inferior parietal lobe. Right hemispheric lesions in the inferior parietal lobe were coupled with impaired familiar voice-identity recognition (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989) (Fig. 1B).

Figure 1. Model of voice-identity processing and overview of neuroanatomical representations of voice-identity recognition.



A. Voice-identity processing is conceived as a multistage process, which comprises (i) perceptual voice-identity processing, (ii) voice-identity recognition (feeling of familiarity), and (iii) semantic associations to a recognised voice (Roswandowitz *et al.*, under review-a). B. Divergent findings on the neural representation of voice-identity recognition. Lesion studies suggest the right inferior temporal lobe and neuroimaging studies the temporal lobe being crucially involved in voice-identity recognition.

Current neuroanatomical models of voice-identity processing, however, do not prominently feature the inferior parietal lobe (Belin *et al.*, 2004; Blank *et al.*, 2014). In contrast, they postulate a central role of the temporal lobe for voice-identity recognition (Fig. 1B). The temporal lobe houses the so-called temporal voice areas (TVAs) that are located along the STG/S. Neuroimaging studies in healthy humans as well as non-human primates and dogs have found that these regions have a strong preference for voices (Belin *et al.*, 2000; von Kriegstein and Giraud, 2006; Petkov *et al.*, 2009; Perrodin *et al.*, 2011; Andics *et al.*, 2014; Pernet *et al.*, 2015; for review see Perrodin *et al.*, 2015). Of these regions, especially right hemispheric superior temporal gyrus/ sulcus (STG/S) regions, extending to the middle temporal gyrus (MTG), are responsive to identity recognition of humans voices (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Formisano *et al.*, 2008; for reviews see Belin *et al.*, 2011; Mathias and von Kriegstein, 2014).

For the perceptual voice-identity analysis, in contrast, neuroimaging and lesion studies consistently report the bilateral temporal lobe (Van Lancker and Kreiman, 1987; Van Lancker *et*

al., 1988; Van Lancker *et al.*, 1989). Within the temporal lobe, neuroimaging studies found distinct sub-regions supporting either the perceptual voice-identity analysis, i.e. posterior STG/S, parts of Heschl's gyrus and planum temporale (Belin *et al.*, 2000; von Kriegstein *et al.*, 2006; Warren *et al.*, 2006; Formisano *et al.*, 2008; Bonte *et al.*, 2014) or voice-identity recognition, i.e. mid to anterior STG/S (Belin and Zatorre, 2003; Andics *et al.*, 2010).

The central goal of the present study was to work towards resolving the discrepancy between on one hand the temporal lobe focus in current standard neuroanatomical voice-identity processing models and neuroimaging studies and on the other hand the consistent findings in lesion studies that the inferior parietal lobe is crucial for voice-identity recognition. To do this is important, because neuroimaging studies with neurotypical participants can give predictions about the potential role of brain regions for behaviour. However, only studies with temporary functional or acquired lesions can give the exquisitely valuable insight about the causal role of these regions for voice-identity recognition behaviour.

In previous case studies on acquired phonagnosia, lesion mapping was based on overlay representations of lesions based on computer tomography (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989; Neuner and Schweinberger, 2000). Since these studies were published, magnetic resonance imaging (MRI) with high spatial resolution has become available in addition to complex statistical analysis procedures such as voxel based morphometry (VBM, Ashburner and Friston, 2000) or voxel-based lesion symptom mapping (VLSM, Bates *et al.*, 2003).

Here, we used VLSM on patients with acquired brain lesions to test whether lesions in parietal and/or temporal lobe are associated with selectively impaired voice-identity recognition abilities. The patient sample included patients with a unilateral brain lesion. Patients with severe aphasia or severe cognitive impairments were excluded. We tested the patients on a comprehensive behavioural test battery on voice-identity recognition. In this test battery, we evaluated the patients' abilities to recognise newly learned unfamiliar and familiar voices as well as the abilities to process acoustic voice features, which are important for voice-identity recognition (vocal pitch, vocal timbre; Lavner *et al.*, 2000; Gaudrain *et al.*, 2009). To control for

visual person recognition, the test battery also included assessments of face-identity recognition abilities. All patients additionally took part in a neuropsychological assessment and a pure-tone audiometry. We used the behavioural scores to correlate them with the lesion data based on structural brain scans. Besides addressing our main research question, the test design also allowed to test whether recognition of voices with different amount of familiarity are represented in the same or different neuroanatomical structures.

Materials and methods

Participants

We recruited 70 patients with unilateral lesions at the Clinic for Cognitive Neurology, University-Hospital Leipzig, Germany. Exclusion criteria were severe aphasia, moderate to severe cognitive impairment, and a diagnosis of psychiatric disorders such as dementia, schizophrenia, personality disorder or depression. All patients gave their written informed consent prior to testing. Data were collected in accordance with the Declaration of Helsinki and of the Ethics Committee of the University of Leipzig. All patients received payment for their participation. From the 70 patients, we excluded 12 patients from the analysis: For 11 patients, we stopped behavioural testing because the tasks were too demanding. For one patient, neuroimaging data could not be obtained. In total, we included 58 patients into the VLSM analysis (31 female, 57 right-handed, Oldfield, 1971) (Fig. 1, Table 1, Supplementary Table 1). 31 patients had a lesion in the right hemisphere and 27 patients in the left hemisphere. Lesion types included ischemic stroke (*n* = 34), traumatic brain injury (*n* = 7), intracerebral bleeding (*n* = 6), subarachnoid haemorrhage (*n* = 6) and tumour excision (*n* = 4). 12 of the 58 patients had a mild to moderate aphasia (Supplementary Table 1). 35 of the patients had a diagnosis of mild and six patients of a moderate cognitive disorder (Supplementary Table 1). Irrespective of aphasia and cognitive disorder diagnoses, all patients had good verbal communication skills and all patients stated that experimental tasks were all feasible. An overview of the patient demographics and the results of the neuropsychological assessment can be found in Table 1. Individual patient characteristics are provided in Supplementary Table 1.

Experimental procedures

General procedure

The complete testing procedure included an audiogram, a neuropsychological assessment, a paper-based questionnaire, and a behavioural test battery (i.e. computer tests on voice- and face recognition). All computer tests were carried out on a desktop computer. Participants were comfortably seated facing a 21-inch monitor that displayed the visual stimuli. Auditory stimuli were presented via Headphones (Sennheiser HD 280 pro, Wennebostel, Germany). The sound level was individually adjusted to a comfortable sound pressure level for each participant. The participants' responses were recorded via a keyboard. To ensure that all participants understood the tasks, the experimenter gave oral instructions in addition to written instructions prior to each test. The audiogram and the computer tests were carried out in a sound-attenuated chamber. Patients were tested in two or three sessions on separate days. Patients could request breaks, whenever they wanted. Experiment completion over the separate days including breaks took approximately 5 hours.

Audiometry

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

Neuropsychological assessment

The neuropsychological assessment contained tests on attention (Alertness as a subtest of the Test of Attentional Performance (TAP) Version 2.2, Zimmermann and Fimm, 2009), on auditory and visual-spatial working memory (Digit and Spatial Span as subtests of the WMS-R, Härting *et al.*, 2000), on 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). For each test, we transformed the raw test scores into age- and education-corrected percentile ranks based on task-specific normalised reference values obtained from healthy participants. Only for one of the scores (i.e. memory score of the face-name learning test) we report percent correct values because no normalised reference values were available (Table 1).

Questionnaire on person recognition

We developed a paper-based questionnaire to assess abilities in everyday person-recognition situations. Participants were asked to rate their abilities on general person-recognition abilities and on voice and face recognition on a scale from 1 (excellent) to 5 (very poor). Further, we asked participants about the cues they use for person recognition (e.g., person as a whole, face, voice, clothing, or posture). 44 out of the 58 patients rated each questionnaire item once for the time before and once for the time after lesion onset. An English version of the questionnaire is available online:

http://kriegstein.cbs.mpg.de/questionnaire /questionnaire_patients.pdf. The other 14 patients filled-in an older version of the questionnaire, in which they were asked to provide ratings for their abilities only after lesion onset.

Behavioural test battery

We used a behavioural test battery for assessing voice-identity recognition abilities. This test battery has been 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).

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

In the voice-name and voice-face test participants learned and recognised previously unfamiliar voices. Unfamiliar voices were either associated 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 14 native speakers of standard German (eight female, six male, age range = 21 - 32 years). We instructed all speakers to read the sentences with a normal speech rate and with an emotionally neutral intonation. Each speaker read 41 five-word declarative sentences (about 2 sec long), 5 two-word declarative sentences (about 0.7 sec long), and 5 five-word interrogative sentences (about 2 sec long). This resulted in a total set of 714 sentences. High quality auditory recordings were taken in a soundproof recording chamber with a condenser microphone (Rode NT 55 MP; USB Sound Interface: Fast Track MK2, M-Audio,

US; 44.1 kHz sampling rate, 16 bit resolution) and Audacity software (version 1.3.5. beta (http://audacity.sourceforge.net). They were recorded using Sound Studio 3 (Felt Tip Inc., NY, USA), post-processed using Audacity (version 1.3.5. beta (http://audacity.sourceforge.net) and Matlab (version 8.1, The MathWorks, Inc., MA, USA), and were normalized for peak amplitude using PRAAT (Boersma and Weenink, 2005). The visual stimuli comprised photographic images of the speakers' faces. Images were recorded with a digital video camera (Legria HF S10 HD-Camcorder, Canon Inc., Japan). They were all taken under the same lighting conditions in front of a black background. The speakers' faces were visible from the chin to the hairline with a neutral expression. No face contained salient visual features such as beards, piercings or glasses. 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 learning test were exactly the same, although different stimuli and different speakers were used in both tests. The tests contained a 'female-voices' block and a 'male-voices' block, which were identical in structure. The blocks were structured into four learning and four interim-testing phases, presented in alternating order, plus an additional final testing phase. During learning, participants heard the five-word declarative sentences spoken by three speakers. Each sentence presentation was either accompanied by a name (voice-name test) or a face (voice-face test). Participants had the task to learn these associations. The learning phases differed in the number of sentences that were presented: ten sentences per speaker during the first and second learning phase and three sentences per speaker in the third and fourth learning phases. The presentation order of the speakers was randomized. During the interim-testing phases, the participants listened to the auditory stimuli and performed a three-alternative forced choice task in which they selected either the name (voice-name test) or the face (voice-face test) associated with the voice. In the first and the second testing phases, participants received feedback about their decision, and the correct voice-name/voice-face pair was presented again. Each interim-testing phase included five sentences per speaker. To avoid prosody-driven identity processing, we used different types of sentences for the testing phases: Five-word declarative sentences for the first and

second phase, five-word interrogative sentences in the third phase, and two-word declarative sentences in the fourth phase. Each test, i.e. voice-name and voice-face test, took approximately 40 minutes to complete. There was a break between both tests.

As a measure of newly-learned voice-recognition performance, we calculated a mean percent correct score of the correct trials over all five testing phases of the female and male block, respectively for the voice-name and voice-face test.

Familiar voice recognition: Famous voice test

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

Stimuli and presentation software

The auditory stimulus set contained voice samples of famous (*n* = 42, see Supplementary methods, famous voice test) and non-famous (*n* = 20) German speakers. We extracted the voice samples from open-access high quality audio files available on public radio and television websites. Each sample was cut to five seconds duration. The famous voice samples comprised voices from 21 media personalities, eight politicians, seven actors and four musicians. In a pilot study, a group of 10 individuals without voice-recognition deficits rated the familiarity of a larger stimulus set (*n* = 56) on a scale from 0 (completely unfamiliar) to 5 (highly familiar). Samples rated with an average of three or higher (*n* = 42) were included in the final stimulus set. The files were edited using Audacity (version 1.3.5. beta (http://audacity.sourceforge.net)) and peak amplitude was scaled using PRAAT (Boersma and Weenink, 2005). The semantic content of the auditory samples provided no information about the celebrities' identity or profession. The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure and analysis

At the beginning of the test, participants were asked to estimate their weekly exposure to television and radio (in hours). Subsequently, the famous and non-famous voice samples were presented. After each sound sample, participants were asked to categorize the voices as familiar or unfamiliar (i.e. familiarity decision). If on a given trial the voice sample was categorized as unfamiliar, the next voice sample was presented. If participants categorised the voice as familiar, irrespective of whether the voice was famous or non-famous, they afterwards performed a four-alternative forced choice task ('semantic association'). A choice of three celebrities (name and face picture) and a question mark was presented on the screen. One of the three celebrities was the same identity as the voice sample. Participants were instructed to choose the question mark if they had thought that the voice sample was belonging to a different celebrity than the ones presented on the screen. Though, for each famous voice trials there was a matching voice-name/face pair. The response options differed from the previous test design (Roswandowitz *et al.*, 2014), in which we asked participants to type in person-specific knowledge in an open-response format. The reasoning behind the new task design was to avoid possible biases due to mild or moderate aphasia.

Before starting the testing, one practice trial was introduced to familiarise participants with the task.

At the end of the test, we assessed each participant's familiarity with the celebrities' identities. To do that, participants were presented with each famous person's written name and face (in a different order than in the main test). In addition, we included 15 celebrities that had not been presented in the main test to get familiarity ratings that are, as much as possible, unbiased by the recognition performance. First, participants were asked to indicate whether the famous person was familiar or unfamiliar to them. If the person was rated as familiar, three questions followed: (i) How often have you heard the voice? (never/ rarely/ sometimes/ often/ very often) (ii) How good do you think you would recognise the person's voice? (not at all/ poor/ good/ very good) (iii) How good do you think you would recognise the person's face? (not at all/ poor/ good/ very good). The complete test took about 45 minutes with a break between the

voice-recognition test (i.e. 30 minutes) and the follow-up survey on celebrity familiarity (i.e. 15 minutes).

Based on their personal familiarity ratings obtained from the survey, we individually re-classified all of the famous voices (*n* = 42) into famous and non-famous voice categories. The celebrities who (i) were known to the participants and (ii) had been at least sometimes listened to were categorised as subjectively familiar persons. All other celebrities were categorized as unfamiliar. The non-famous voices (*n* = 20) were all classified as non-famous. This procedure allowed us to determine whether the participants correctly or incorrectly categorized the voices of subjectively familiar celebrities. We analysed the data by applying detection theory (Macmillan and Creelman, 2004). We computed indices of sensitivity (d-prime, *d'*) of voice familiarity judgment assuming the yes-no decision model. Each famous voice sample correctly classified as familiar was considered a 'hit', and each famous voice sample classified as unfamiliar was considered a 'false alarm' (i.e. familiarity decision).

In addition, we computed the proportion of those familiar voices that were correctly matched with the name and face of the respective celebrity (i.e. semantic association) and those that were not correctly matched or for which the question mark was chosen.

Acoustic voice processing: Vocal-pitch and vocal-timbre test

Using an adaptive tracking procedure (Kaernbach, 1991), we measured the individual threshold to discriminate between two subsequent vowels, modulated in pitch (vocal-pitch test) or timbre (vocal-timbre test). For details see Supplementary methods.

Visual control tests

CFMT

With the CFMT (Duchaine and Nakayama, 2006), we investigated the ability to recognise unfamiliar faces. Participants first learned six male faces. They were then asked to recognise one of the learned faces out of three presented faces (one learned, two unfamiliar). The CFMT comprises three different test sections: same images, novel images, and novel images added with Gaussian noise. There was a total of 72 items. The test took approximately 15 minutes.

Face-name test

In the face-name test, participants learned and recognised previously unfamiliar faces and their names. 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 they were asked to select the corresponding name from six alternatives. For details see Supplementary methods.

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, Erlangen, Germany) including 3D T1-weighted- (1mm^3 isotropic voxels), and FLAIR-images. The lesions were manually delineated in all three planes (axial, coronal, sagittal) on each slice of the T1-images using MRIcron (Rorden and Brett, 2000). The FLAIR-images served as a reference. Lesion delineation was performed by an experienced neurologist (HO). In addition, lesion delineation was checked by a second experienced neuroscientist with medical training (KVK). Both were blind to the individual patient's performance in the behavioural test-battery. T1 images were then 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).

Voxel-based lesion symptom mapping analysis

We performed voxel-based lesion symptom mapping (VLSM, Bates *et al.*, 2003) to identify systematic relationships between lesioned brain regions and behavioural measures. We used VLSM 2.55 (http://www.neuroling.arizona.edu/resources.html) implemented in Matlab (version 8.2, The MathWorks, Inc., MA, USA). For each behavioural test, the spatially normalised lesion maps and corresponding behavioural scores were used to create *t*-maps. On a voxel-by-voxel basis, the VLSM analyses run a general linear model comparing performances on every measure in patients with a lesion versus without a lesion in the respective voxel. Thus, lesion status of the voxel (0 = intact, 1 = lesioned) is the independent variable and the behavioural measures

the dependent variables. We applied several general linear models, i.e. with the behavioural measures of each test separately and with composite scores as dependent variables. To compute composite scores of measures with different units (i.e. percent correct scores of unfamiliar voice tests and d' prime scores of the familiarity decision of the famous voice test), we z-transformed them (for more details see Results). For each general linear model, we calculated an additional analysis with hearing level (mean over both ears) and lesion volume as covariates to account for a possible influence of these factors. For the VLSM analyses, we included only voxels in which at least 3 patients had a lesion.

Significance threshold

Statistical maps were thresholded at voxelwise $p < 0.01$. 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 ROI analyses. For the behavioural analyses, we applied a significance threshold of $p < 0.05$ and applied Bonferroni-correction.

Region of interest definition

We created a ROI map including the bilateral temporal lobe and the right inferior parietal lobe to investigate brain structures associated with voice-identity processing. The ROI 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 temporal lobes from the MNI Structural Atlas (Mazziotta *et al.*, 2001) and for the right inferior parietal lobe from the Juelich Histological Atlas (Caspers *et al.*, 2013). The inferior parietal lobe map contained the areas PF, PFm, and Pga. The resulting map covered parietal regions that have been previously associated with voice-identity recognition deficits i.e., the supramarginal gyrus (SMG), and the 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 is relatively unspecific as it also includes regions that have not been reported to be critically involved in voice-identity processing (i.e., lateral occipital cortex, postcentral gyrus, postcentral gyrus, precuneus, cingulate gyrus), and (ii) a temporal lobe map is not provided in the Juelich

Histological atlas. Based on visual inspection, we chose a threshold of 10% to restrict maps to anatomically meaningful brain regions. The final ROI map was a union of the bilateral temporal lobe and the right inferior parietal lobe maps.

VLSM results reporting

For the VLSM results, we report coordinates for the clusters' centre of mass.

For anatomical labelling we used three atlases provided by the FSL Anatomy Toolbox: (i) the Juelich Histological Atlas (Eickhoff *et al.*, 2005) for sub-classification 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

Results on the behavioural test battery, neuropsychological assessment, and audiometry are displayed in Table 1. For all of the voice and face recognition tests, patients performed significantly above chance level (Table 1). We used Pearson correlations to assess whether the test performance was influenced by demographic variables (age, time since lesion onset, education, and lesion volume), hearing level, or neuropsychological status. No correlation reached significance after correcting for multiple comparisons ($p < 0.05$, after Bonferroni correction for 99 tests $p < 0.0005$). Correlation results at an uncorrected level are reported in Supplementary Results.

Table 1: Demographic details, neuropsychological, and behavioural measures.

	n	Mean	SD	Min	Max
Demographic details					
Age	58	47.95	11.59	22.00	68.00
Time since onset [month]	58	46.02	51.51	2.00	273.00
Education [years]	58	10.50	1.23	8.00	12.00
Hearing level [dB]	58	18.11	8.45	3.34	39.17
Neuropsychological tests					
Digit span [PR]	58	32.15	22.16	2.00	90
Spatial Span [PR]	58	39.35	24.95	4.50	95
TAP (alertness) [PR]	54	32.77	23.64	1.00	86
German Vocabulary Test [PR]	54	54.51	20.61	8.10	91.90
Face-name learning	53				
cued + sum [PR]		61.34	34.36	0.00	100
memory [%]		87.56	22.04	3.00	120
Behavioural test battery					
Newly-learned voice recognition					
Newly-learned voice tests [%]	43	59.97	10.30	46.00	87.22
Voice-name test [%]	58	56.14	12.91	31.67	90.00
Voice-face test [%]	43	64.08	11.66	45.34	93.33
Familiar voice recognition					
Famous voice test [z]	57	0.00	0.73	-1.28	1.84
Familiarity judgment [d']	57	1.11	0.84	-1.67	3.65
Semantic association [%]	57	69.60	18.81	25.00	100.00
Acoustical voice processing					
Vocal-pitch test [cent]	58	125.17	66.05	18.28	249.96
Vocal-timbre test [SER]	55	9.73	4.62	1.66	17.81
Visual control tests					
CFMT [%]	57	66.79	15.52	34.72	93.06
Face-name test [%]	43	60.54	19.15	20.67	96.67

The table displays mean values, standard deviations, minimum (Min) and maximum (Max) values on demographical, neuropsychological measures, and on each behavioural test respectively for all patients who completed the test (*n*). Hearing levels are averaged over both ears. PR = percentage rank, TAP = Test of attentional performance, SER = spatial envelop ratio.

Comparing behavioural performance in patients with left and right sided lesions has a long tradition in voice-recognition research (Assal *et al.*, 1981; Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1989; Lang *et al.*, 2009) and we therefore also compared voice-recognition performance in these two groups. First, we tested if both groups differed in their voice-recognition performance over both familiarities (i.e. composite score of z-transformed means of the newly-learned (i.e. mean of voice-name and voice-face test) and of the familiar voice tests (i.e. mean of familiarity decision and semantic association of the famous voice test). This analysis revealed comparable group performance in patients having a right- and left-hemispheric lesion ($Z = 0.957$, $p = 0.339$). Next, we separately compared group performances of the newly-learned and familiar voices. For the composite score of the newly-learned voice recognition tests, the group of patients with right-hemispheric lesions performed significantly worse than the group of patients with left-hemispheric lesions (composite score of voice-name and voice-face test: $Z = -2.28$, $p = 0.022$, Bonferroni correction for the two voice-recognition measures, i.e. composite score newly-learned voice recognition and composite score of famous voice test, $p = 0.025$). The familiar voice recognition performance was comparable in both lesion groups (composite score of familiarity decision and semantic association: $Z = 1.08$, $p < 0.303$. For completeness, we also checked group differences for the other behavioural measures at an uncorrected level. Of these tests, only the voice-face test revealed a significant group difference, i.e. worse performance in patient group with right-hemispheric lesions compared to the left-hemispheric patient group (see Supplementary Table 2).

Questionnaire results

The results of the person-recognition questionnaire are reported in Supplementary results, and Supplementary Table 3,4.

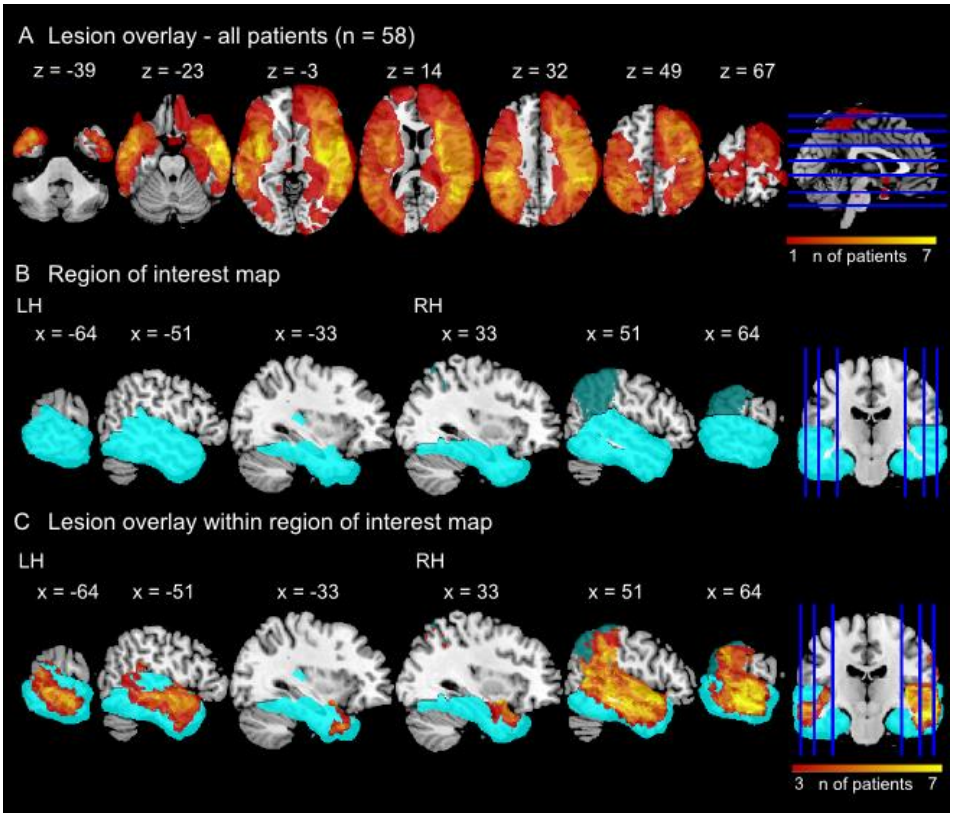
VLSM results

Lesion coverage

Figure 2 displays the lesion overlay for all 58 patients (Fig. 2A). We checked whether lesions in the temporal lobe covered those parts that have been previously characterised as voice-identity sensitive in neuroimaging studies. In the same vein, we also inspected lesion coverage in the inferior parietal lobe that has been previously described as causing voice-identity recognition impairments in patient case reports.

To do that, we checked if the anatomical temporal lobe map (Fig. 2B) covered all reported voice-sensitive statistical maxima found in neuroimaging studies. This was the case for 59 of 60 reported statistical maxima (Belin *et al.*, 2000; Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; von Kriegstein *et al.*, 2005; Warren *et al.*, 2006; Blank *et al.*, 2011; Blank *et al.*, 2014; Bonte *et al.*, 2014; Roswandowitz *et al.*, accepted). We further checked if areas with a-priori anatomical hypotheses within the temporal lobe map were covered by lesions in more than 3 patients. This was the case for all the voice-sensitive regions, i.e. bilateral STG, MTG, Heschl's gyrus and planum temporale. Next, we checked if areas with a-priori anatomical hypotheses of the right inferior parietal lobe map were covered by lesions in more than 3 patients (Fig. 2B). The angular and supramarginal gyrus were well covered (Fig. 2C). For an overview of ROI structures that were not covered by lesions see Supplementary Table 4.

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. Coloured areas are lesioned in at least one patient. Voxels lesioned in seven or more patients are depicted in yellow. Coordinates refer to MNI space. **B.** Region of interest map. Anatomical region of interest map covered the bilateral temporal lobe (i.e. cyan) and the right inferior parietal lobe (i.e. turquoise). **C.** Lesion overlay map of all patients within the region of interest map. Coloured areas are lesioned in at least three patients. Voxels lesioned in seven or more patients are depicted in yellow.

Voice recognition

First, we investigated which lesion locations were associated with deficient voice recognition irrespective of voice familiarity. To do that, 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 voice recognition (familiarity decision and semantic association of the famous voice test) as the dependent variable. There was a statistical brain-behaviour association in a large right temporal lobe cluster with its centre of mass in the planum temporale extending to mid/posterior STG and the inferior parietal lobe i.e. the SMG (Fig. 3A, Table 2). Next, we tested whether our findings remained significant when we control for

individual hearing abilities and lesion volumes. When adding both factors as covariates into the analysis, a right temporal lobe cluster with its centre in the posterior STG remained significant (Fig. 3A, Table 2). Notably, this cluster again extended to the right inferior parietal lobe.

Voice recognition controlled for face recognition

We next tested whether the statistical brain-behaviour association for voice recognition in the right temporal and inferior parietal lobe was selective to vocal person recognition. To do that, we controlled for facial person-recognition abilities by entered the z-transformed composite score of the CFMT and face-name test performance as a covariate into the analysis. This allows the analysis of variance in voice recognition (predictor variable) while minimising face recognition deficits as a source of variance (covariate) (Baldo *et al.*, 2013; Rogalsky *et al.*, 2015; Binder *et al.*, 2016). This analysis revealed a significant brain-behaviour association in the right temporal lobe (Fig. 3B). The cluster's centre of mass was located in the planum temporale (Table 2). After adding hearing level and lesion volume as covariates, the same right temporal lobe cluster remained significant. There were no further significant clusters at a whole-brain level for any of the analyses.

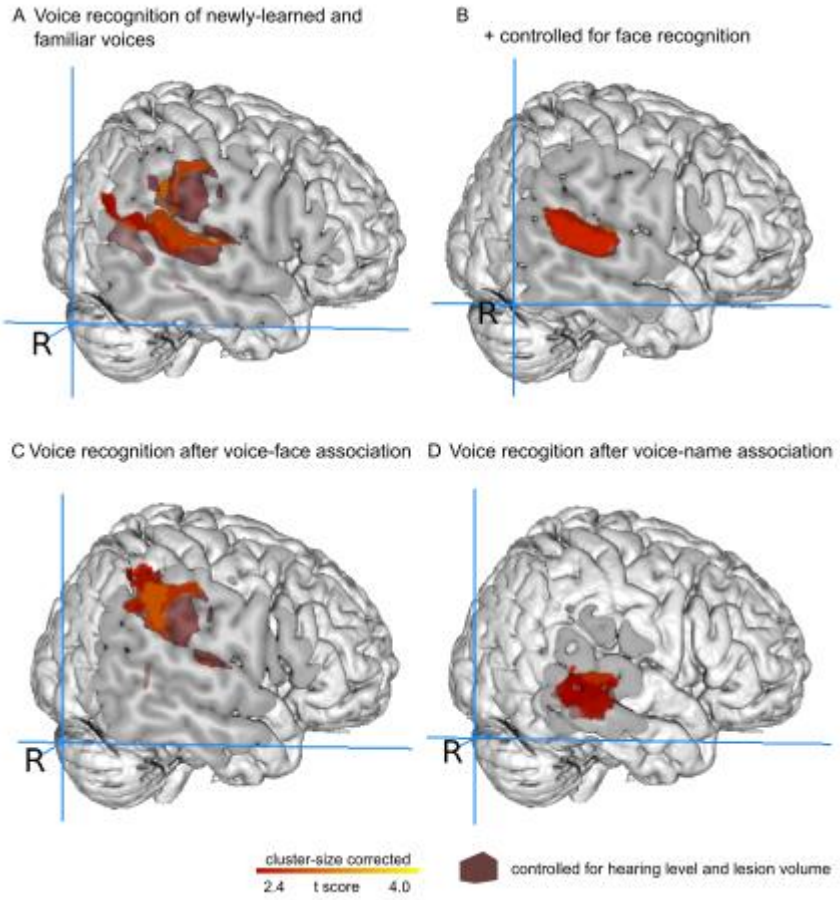
The results suggested that both the temporal and inferior parietal lobe were critical for voice recognition, but that the temporal lobe might be more specifically involved in voice recognition, while the inferior parietal lobe might be rather involved in multimodal person recognition that also involves faces.

Voice recognition after voice-face and voice-name association

To explicitly assess the role of facial information during voice recognition in the inferior parietal lobe, we separately looked at brain-behaviour associations for identity recognition of voices that were associated with faces (i.e. voice-face test) and voices that were associated with names (i.e. voice-name test). For the voice-face test, a lesion cluster associated with deficient performance was located in the right inferior parietal lobe (i.e. parietal operculum cortex, Fig. 3C, Table 3) extending marginally into the temporal lobe. This cluster remained significant after adding hearing level and lesion volume into the analysis (Fig. 3C, Table 2). For the voice-name test, a lesion cluster associated with deficits in performance was located in the right temporal lobe (i.e. posterior MTG, Fig. 3D, Table 2) with a small extension into the inferior parietal lobe.

The effect in the temporal lobe remained significant after adding hearing level and lesion volume as covariates. Please note, although the significant peak clusters of both tests were located in distinct lobes, there was an overlap of lesions in the right temporal lobe (i.e. posterior STG) (controlled for hearing level and lesion volume).

Figure 3. Lesions in the temporal and inferior parietal lobe associated with difficulties in voice recognition.



A. VLSM results for the composite score of newly-learned and familiar voice recognition (*n* = 42). **B.** VLSM results for voice-recognition controlled for performance in face recognition (i.e., composite score of CFMT and face-name test, *n* = 40). **C.** VLSM results of the voice-face test (*n* = 43). **D.** Results of the voice-name test (*n* = 58). VLSM results that are controlled for hearing level and lesion size are overlaid as red surfaces on each image. Analyses were restricted to the bilateral temporal lobe and right inferior parietal lobe. All voxels shown exceeded the critical threshold for significance ($p < 0.01$ cluster-size corrected, 1000 permutations).

Table 2. Overview of lesion-cluster coordinates associated with voice recognition.

Region	Voice recognition (newly-learned and familiar voices)						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right temporal lobe												
Planum temporale	57	-34	16	3.96	0.005	12460						
Posterior STG							56	-37	13	3.51	0.004	15656

	Voice recognition Face recognition controlled						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right temporal lobe												
Planum temporale	59	-30	12	3.38	0.038	5442	59	-30	12	3.26	0.038	5442

	Voice-face test						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right parietal lobe												
Parietal operculum cortex	59	-33	32	3.36	0.003	6921	57	-34	28	3.30	0.013	8499

	Voice-name test						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right temporal lobe												
posterior MTG	62	-31	0	3.03	0.030	3532	60	-35	-1	3.73	0.023	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). Results are reported at $p < 0.01$ cluster-size corrected, 1000 permutations. STG = superior temporal gyrus, MTG = middle temporal gyrus, HL = hearing level, LV = lesion volume

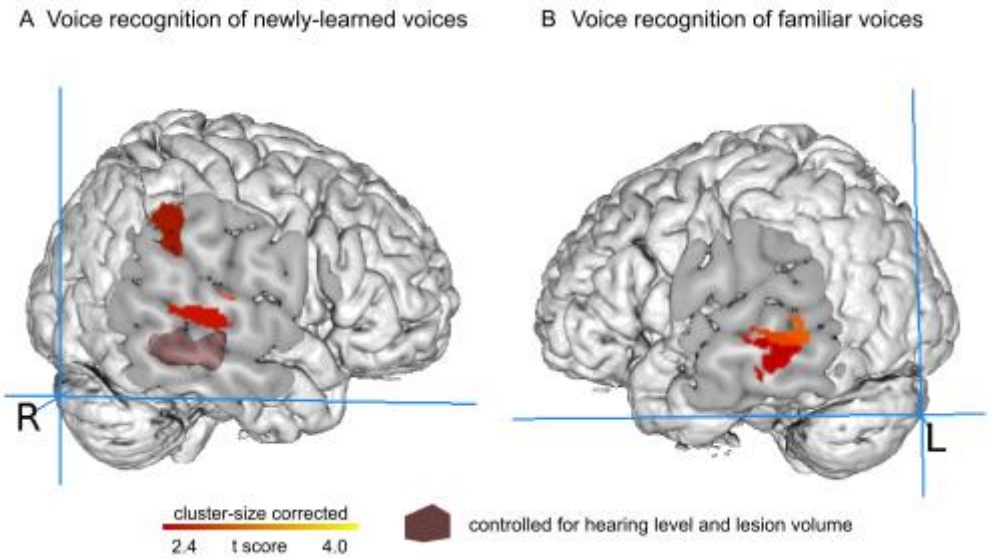
Newly-learned and familiar voice recognition

Lesion studies have shown that impaired identity processing of different voice familiarities are related to different brain lesions (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989; Hailstone *et al.*, 2011). However, in those studies different test designs have been used, i.e. discrimination/matching tasks for unfamiliar voice-identity processing and recognition tasks for familiar voice-identity processing. We here tested whether also different lesion clusters are associated with impaired identity processing of voices having different familiarities when using the same test design, i.e. voice-recognition test.

To assess lesion clusters associated with newly-learned voice recognition, we performed a VLSM analysis including the composite score of the voice-name and the voice-face test performance as the dependent variable. We found a significant brain-behaviour association in both the inferior parietal and temporal lobe of the right hemisphere (Fig. 4A, Table 3). In the right inferior parietal lobe, a cluster in the posterior SMG reached significance and in the right temporal lobe four clusters reached significance: the posterior STG, the posterior MTG, the planum temporale and Heschl's gyrus. Lesion clusters in the temporal (i.e. posterior MTG) and inferior parietal lobe (i.e. posterior SMG) remained significant after controlling for hearing level and lesion volume (Fig. 4A, Table 3).

Next, we assessed the brain-behaviour association for familiar voice recognition. We included the z-transformed composite score of familiarity decision and semantic association of the famous voice test as a dependent variable. There was a significant brain-behaviour association in the left temporal lobe, i.e. posterior MTG (Fig. 4B), Table 3). After adding hearing level and lesions volume as covariates no cluster remained significant. 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, Supplementary Table 6).

Figure 4. Lesions associated with identity recognition of newly-learned and familiar voices.



A. VLSM results for newly-learned voice recognition (composite score of voice-name and voice-face test, *n* = 43 patients). VLSM results that are controlled for hearing level and lesion size are overlaid as red surfaces. **B.** VLSM results for familiar voice recognition (composite score of the familiarity decision and semantic association score of the famous voice test, *n* = 57 patients). All voxels shown exceeded the critical threshold for significance ($p < 0.01$ cluster-size corrected, 1000 permutations). Analysis is 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 recognition.

Region	Newly-learned voice recognition (voice-name and voice-face test)						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right parietal lobe												
posterior SMG	58	-37	32	3.25	0.002	1971	57	-39	36	3.02	0.031	2133
Right temporal lobe												
posterior STG	62	-27	3	2.69	0.003	1116						
posterior MTG	48	-37	-3	2.62	0.018	90	58	-31	-5	3.24	0.019	4087
Planum temporale	59	-17	8	2.77	0.023	59						
Heschl's gyrus (Te 1.1, Te 1.0)	44	-23	10	2.50	0.036	17						
	Familiar voice recognition (familiarity and semantic association)						+ HL and LV					
Region	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Left temporal lobe												
posterior MTG	-55	-33	-8	2.99	0.04	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). Results are reported at $p < 0.01$ cluster-size corrected, 1000 permutations. SMG = supramarginal gyrus, STG = superior temporal gyrus, MTG = middle temporal gyrus. HL = hearing level, LV = lesion volume

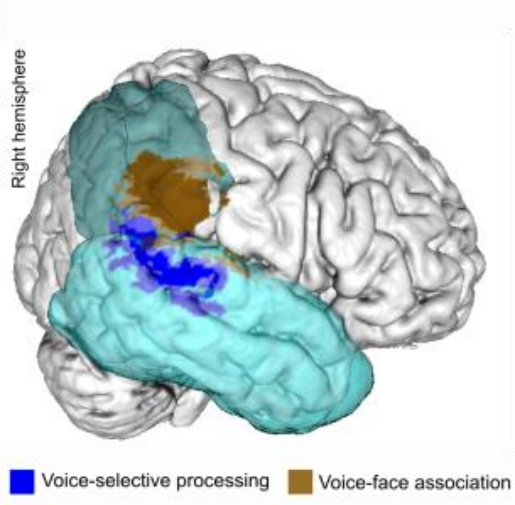
Acoustical voice processing

The analyses of the brain-behaviour association on acoustical voice processing (i.e. vocal-pitch and vocal-timbre test) revealed no significant results.

Discussion

We here used voxel-based lesion symptom mapping to investigate the contribution of temporal and inferior parietal lobe lesions to difficulties in voice-identity recognition in patients with acquired brain lesions. There were two key findings. First, lesions in both the temporal and the right inferior parietal lobe were associated with voice-recognition impairments. However, if we controlled for face recognition difficulties, only lesions in the right temporal lobe were associated with impaired voice recognition. Second, lesions in the right inferior parietal lobe were particularly associated with difficulties in voice recognition, if the voices had been learned together with the corresponding face, but not to the same extend if they had been learned together with a name. For a schematic overview of the key brain-behaviour associations see Fig. 5. In addition, identity recognition of different voice familiarities was accomplished in distinct brain structures. Difficulties in recognising newly-learned voices were associated with lesions in the right hemisphere and difficulties with familiar voice recognition with left-hemispheric lesions. The findings provide an important step forward in resolving the current discrepancy between neuroimaging and lesion findings, because they (i) qualify the contributions of the inferior parietal lobe in voice-identity recognition and causally confirm (ii) the crucial contributions of the right temporal lobe to successfully recognise voice identities.

Figure 5. Overview of lesions associated with voice-recognition deficits within the temporal (cyan map) and inferior parietal lobe (turquoise map).



Selective voice recognition in the temporal lobe

Our findings are in line with the numerous investigations on neurotypical participants that repeatedly showed voice sensitivity in the temporal lobe. These responses have been found for different test designs (e.g. vocal-sound listening, voice-identity recognition, and voice discrimination) and different voice familiarities (e.g. unfamiliar, newly-learned, and personally familiar voices) (Belin *et al.*, 2000; Shah *et al.*, 2001; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Warren *et al.*, 2006; Bestelmeyer *et al.*, 2011). Also, voice sensitivity in the temporal lobe is supported by different imaging methods such as fMRI (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003), EEG (De Lucia *et al.*, 2010), MEG (Renvall *et al.*, 2012; Capilla *et al.*, 2013; Schall *et al.*, 2015), as well as one neurostimulation study (Bestelmeyer *et al.*, 2011). The importance of the temporal lobe is also highlighted in traditional lesions studies. There the temporal lobe was linked to the perceptual analysis of unfamiliar voices (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989).

Here we advance previous neuroimaging as well as clinical findings, because we prove a causal contribution of the temporal lobe not only to the perceptual voice analysis, as previously shown, but also for the recognition of voice identities. Further, we precisely showed which lesioned sub-structures of the right temporal lobe cause reduced abilities in voice recognition and that these are relatively selective: They were independent from abilities for face recognition and they were present in patients with relatively intact language skills (Fig. 2B).

Current neuroscientific models suggest that selective voice-identity processing, including perceptual voice analysis and voice-identity recognition, is processed in the core-voice system in the temporal lobe (Belin and Zatorre, 2003; von Kriegstein and Giraud, 2004; Warren *et al.*, 2006; Andics *et al.*, 2010; Roswandowitz *et al.*, under review). The semantic association to a recognised voice, however, in an extended system, which shares connections to the core-voice system (Hailstone *et al.*, 2011; Gainotti, 2015; Roswandowitz *et al.*, under review). Within the temporal lobe neuroimaging studies have found that posterior STG/STS and auditory regions are involved in the perceptual analysis of complex spectrotemporal voice features that accomplish voice-identity processing, while more anterior regions seem to represent voice

identity (Belin and Zatorre, 2003; von Kriegstein and Giraud, 2004; von Kriegstein *et al.*, 2007; Andics *et al.*, 2010).

The lesion locations in the present study that lead to difficulties in selective voice-identity recognition are located in rather posterior STG/STS and auditory regions (planum temporale and Heschl's gyrus) of the core-voice system.

By using a comprehensive test battery and sophisticated analysis methods, a recent VBM study on voice-identity processing gave the first spatially precise indication that is it the anterior temporal lobe that crucially supports voice-identity recognition (Hailstone *et al.*, 2011). However, lesions in the anterior medial temporal lobe were not exclusively linked to impaired voice-identity recognition; they were rather associated with multi-modal person recognition deficits (i.e. by voice, face, and name). This is consistent with previous reports of associative person-recognition deficits with anterior temporal lobe lesions in neurodegenerative disease (Gainotti *et al.*, 2003; Gainotti *et al.*, 2008; Hailstone *et al.*, 2010).

Also neuroimaging studies have assigned voice-identity representations to anterior regions of the temporal lobe, i.e. anterior STG/S (Belin and Zatorre, 2003; Andics *et al.*, 2010). Importantly, those studies were not conclusive whether neural response in the anterior STG/S was selective to voice-identities or rather of multi-modal nature. Although selective voice-identity processing in the anterior STS has been proven in monkeys (Perrodin *et al.*, 2011), evidence in humans is missing to date.

Integrative voice recognition in the right parietal lobe

Our inferior parietal lobe finding is in line with the previous lesion literature showing that voice-identity recognition impairments occur after right inferior parietal lobe lesions. In particular, the SMG and angular gyrus of the right inferior parietal lobe were involved in famous voice recognition tasks (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). In contrast to these traditional lesion studies, a recent VBM study associated lesions in the right angular gyrus with impaired perceptual voice-identity processing (i.e. unfamiliar speaker change detection task) (Hailstone *et al.*, 2011). In the current study, we found lesions in the SMG being associated with

impaired identity recognition of newly-learned and familiar voices. Taken previous and current lesion findings together, the right SMG was repeatedly associated with voice-identity recognition whereas the angular gyrus was found for both perceptual and identity processing of voices. Based on the consistent finding in the SMG, we argue the SMG might be the most relevant sub-structure of the inferior parietal lobe to accomplish voice-identity recognition.

Interestingly, our findings present several indications that the right inferior parietal lobe is predominantly involved in voice-identity recognition if it is tested via voice-face associations. First, the inferior parietal lobe association with voice recognition was reduced when we controlled for abilities in face recognition (Fig. 2B). Second, although the parietal lobe showed significant association with newly-learned voice recognition in general, this was mostly the case for the voice-face test, but not for the voice-name test (Fig. 3C/D). Previous lesion studies that found right inferior parietal lobe involvement tested patients on how well they associated face/name pairs to a recognised famous voice (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). In the same vein, neuroimaging studies have shown that the right inferior parietal lobe and the adjacent posterior STG region are involved in the integration of person-related voice and face information (von Kriegstein and Giraud, 2006; for review see Campanella and Belin, 2007; Ethofer *et al.*, 2013; Watson *et al.*, 2014). We therefore speculate that the inferior parietal lobe has something to do with associating a face (or potentially also other sensory information) to the voice.

Voice recognition of familiar and newly-learned voices

Within the temporal lobe voice-recognition deficits were associated with lesions in the posterior temporal lobe. However, different hemispheres were involved dependent on the voice familiarity. For newly-learned voices it was the right temporal lobe and for familiar voices it was the left temporal lobe that showed significant associations. Next to the temporal lobe, deficits in newly-learned voices were also associated with lesions in the right inferior parietal lobe. That we found distinct lesion clusters for voices of different familiarities is in line with traditional patient reports (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). These studies used different test designs to assess voice-identity processing, i.e. discrimination tests for unfamiliar voices and recognition tests for familiar

voices and report the bilateral temporal lobe for unfamiliar voice discrimination and the right inferior parietal lobe for familiar voice recognition. In contrast, we here assessed the same test design to investigate voice-identity processing of different familiarities, i.e. voice-recognition tests. The different experimental approaches might explain why we found partly different lesion clusters.

We identified a left-hemispheric lesion association for deficient familiar voice recognition. To our knowledge this has so far neither been reported in previous neuroimaging nor lesion studies. However, neuroimaging studies on voice-identity processing of different familiarities tend to also report left posterior temporal lobe clusters (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Warren *et al.*, 2006). The function of these regions is entirely unclear. On one hand they could represent perceptual processing specific to voices, on the other hand they could also have something to do with association of semantic information (Blank *et al.*, 2014; Liebenthal *et al.*, 2014; Vitali *et al.*, 2015) or representation of names (Blank et al., 2014).

In line with previous reports, we also found the right inferior parietal lobe involved in voice-identity recognition (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). As Van Lancker *et al.* (1989), we argue that familiar voice recognition requires the matching of voice-individuating information to the incoming voice. Assuming voice-face associations in the inferior parietal lobe, we think that this multi-modal association might facilitate the recognition of voices that have been encountered before such as familiar and newly-learned voices. Here, we find the inferior parietal lobe mainly when voices, which were learned with a face, had to be recognised. In contrast, for unfamiliar voice-identity processing the inferior parietal lobe has rarely been reported. We think that this is reasonable as for voices that we hear for the first time individuating voice information such as the facial representation is unlikely to be available.

In summary, our study made important steps towards resolving the discrepancy between neuroimaging and lesion reports on voice-identity recognition, because it shows that (i) the right temporal lobe is crucially involved in voice-identity recognition in general and that (ii) parietal lobe lesions likely lead to voice-identity recognition deficits only if the task requires

voice-face matching. This has important implications for current voice-identity processing models.

First, it supports the central assumption that the temporal lobe is the key structure of the core-voice system. Second, it also opens up two novel lines of research. First how can one integrate the specific nature of voice-face integration in the parietal lobe in person-recognition models? Second, in how far does the level of voice familiarity impacts already on the stage before the voice is recognised as familiar, i.e. during acoustical voice-identity processing?

Authors contributions

KvK conceived the experiment, CR, KvK designed the experiment, CR, HO implemented the design, HO, CR recruited the patients, CR, CK performed the experiment, CR, CK, HO, KVK analysed the data, CR, KvK wrote the paper.

Acknowledgments

We are grateful to all patients for the time and effort taken to participate in the study. We also thank Anne-Katrin Franz for her great help with inviting the patients. We thank Angelika Otto-Thöne for advice on the selection of the neuropsychological tests. We also thank Stephen Wilson for his technical advices supporting the VLSM analysis.

Funding

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

References

Alba-Ferrara L, Weis S, Damjanovic L, Rowett M, Hausmann M. Voice Identity Recognition Failure in Patients With Schizophrenia. The Journal of Nervous and Mental Disease 2012; 200(9): 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. Current biology : CB 2014; 24(5): 574-8.

Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z. Neural mechanisms for voice recognition. NeuroImage 2010; 52(4): 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(3): 839-51.

Assal G, Buttet J, Thuillard F. [Cerebra hemispheres and auditory perception]. Revue medicale de la Suisse romande 1981; 101(3): 177-85.

Assal G, Zander E, Kremin H, Buttet J. [Voice discrimination in patients with cerebral cortical lesions]. Schweizer Archiv fur Neurologie, Neurochirurgie und Psychiatrie = Archives suisses de neurologie, neurochirurgie et de psychiatrie 1976; 119(2): 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(3): 658-67.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT*, et al.* Voxel-based lesion-symptom mapping. Nature neuroscience 2003; 6(5): 448-50.

Belin P, Bestelmeyer PE, Latinus M, Watson R. Understanding voice perception. British journal of psychology 2011; 102(4): 711-25.

Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 2004; 8(3): 129-35.

Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 2003; 14(16): 2105-9.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403(6767): 309-12.

Bestelmeyer PE, Belin P, Grosbras MH. Right temporal TMS impairs voice detection. Current biology : CB 2011; 21(20): R838-9.

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.

Blank H, Anwander A, von Kriegstein K. Direct Structural Connections between Voice- and Face-Recognition Areas. The Journal of neuroscience : the official journal of the Society for Neuroscience 2011; 31(36): 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 2014; 36(1): 324-39.

Boersma P, Weenink D. Praat: doing phonetics by computer (Version 4.3.14). 2005.

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. The Journal of neuroscience : the official journal of the Society for Neuroscience 2014; 34(13): 4548-57.
Boucher J, Lewis V, Collis G. Familiar face and voice matching and recognition in children with autism. Journal of child psychology and psychiatry, and allied disciplines 1998; 39(2): 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(2): 486-500.
Campanella S, Belin P. Integrating face and voice in person perception. Trends Cogn Sci 2007; 11(12): 535-43.
Capilla A, Belin P, Gross J. The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cerebral cortex 2013; 23(6): 1388-95.
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. Cerebral cortex 2013; 23(3): 615-28.
De Lucia M, Clarke S, Murray MM. A temporal hierarchy for conspecific vocalization discrimination in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience 2010; 30(33): 11210-21.
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(3): 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(4): 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(4): 1325-35.
Ethofer T, Bretscher J, Wiethoff S, Bisch J, Schlipf S, Wildgruber D*, et al.* Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. Neuroimage 2013; 76: 45-56.
Fine DR. A life with prosopagnosia. Cognitive neuropsychology 2012; 29(5-6): 354-9.
Formisano E, De Martino F, Bonte M, Goebel R. "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science 2008; 322(5903): 970-3.
Gainotti G. Implications of recent findings for current cognitive models of familiar people recognition. Neuropsychologia 2015; 77: 279-87.
Gainotti G, Barbier A, Marra C. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. Brain : a journal of neurology 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; a journal devoted to the study of the nervous system and behavior 2008; 44(3): 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(1): 123-31.

Gaudrain E, Li S, Ban VS, Patterson RD. The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Interspeech 2009: 10th Annual Conference of the International Speech Communication Association 2009 2009; 1-5: 152-5.
Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia 2010; 48(4): 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(9): 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.
Kaernbach C. Simple adaptive testing with the weighted up-down method. Perception & psychophysics 1991; 49(3): 227-9.
Kimberg DY, Coslett HB, Schwartz MF. Power in Voxel-based lesion-symptom mapping. Journal of cognitive neuroscience 2007; 19(7): 1067-80.
Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG. Voice recognition in aphasic and non-aphasic stroke patients. Journal of neurology 2009; 2009/04/09(Journal Article).
Lavner Y, Gath I, Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun 2000; 30(Journal Article): 9-26.
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. Frontiers in neuroscience 2014; 8: 289.
Macmillan NA, Creelman CD. Detection theory: A user's guide. New York, London: Psychology press; 2004.
Mathias SR, von Kriegstein K. How do we recognise who is speaking? Frontiers in bioscience 2014; 6: 92-109.
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(1412): 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(3): 342-66.
Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971; 9(1): 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 in cognitive sciences 2015; 19(12): 783-96.
Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice cells in the primate temporal lobe. Current biology : CB 2011; 21(16): 1408-15.

Petkov CI, Logothetis NK, Obleser J. Where are the human speech and voice regions, and do other animals have anything like them? The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry 2009; 15(5): 419-29.
Renvall H, Staeren N, Siep N, Esposito F, Jensen O, Formisano E. Of cats and women: temporal dynamics in the right temporoparietal cortex reflect auditory categorical processing of vocalizations. Neuroimage 2012; 62(3): 1877-83.
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(4): 191-200.
Roswandowitz C, Mathias Samuel R, Hintz F, Kreitewolf J, Schelinski S, von Kriegstein K. Two Cases of Selective Developmental Voice-Recognition Impairments. Current Biology 2014; 24(19): 2348-53.
Roswandowitz C, Schelinski S, von Kriegstein K. Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. NeuroImage accepted.
Roswandowitz C, Schelinski S, von Kriegstein K. Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. under review.
Schall S, Kiebel SJ, Maess B, von Kriegstein K. Voice identity recognition: functional division of the right STS and its behavioral relevance. Journal of cognitive neuroscience 2015; 27(2): 280-91.
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(11): 1812-22.
Schelinski S, Roswandowitz C, von Kriegstein K. Voice identity processing in autism spectrum disorder. Autism Res 2016b.
Schmidt KH, Metzler P. Wortschatztest (WST). Hogrefe, Göttingen 1992.
Schuri U, Benz R. Gesichter-Namen-Lerntest: Swets Test Services; 2000.
Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ*, et al.* The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 2001; 124(Pt 4): 804-15.
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(5): 829-34.
Van Lancker DR, Canter JG. Impairment of Voice and Face Recognition in Patients with Hemispheric Damage. Brain Cogn 1982; 1(Journal Article): 185-95.
Van Lancker DR, Cummings JL, Kreiman J, B.H D. Phonagnosia; A dissociation between familiar and unfamiliar voices. Cortex; a journal devoted to the study of the nervous system and behavior 1988; 24(Journal Article): 195-209.
Van Lancker DR, Kreiman J, Cummings J. Voice perception deficits: neuroanatomical correlates of phonagnosia. JClinExpNeuropsychol 1989; 11(5): 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(5): 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 ResCogn Brain Res 2003; 17(1): 48-55.
von Kriegstein K, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 2004; 22(2): 948-55.
von Kriegstein K, Giraud AL. Implicit Multisensory Associations Influence Voice Recognition. PLoSBiol 2006; 4(10).
von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. Journal of cognitive neuroscience 2005; 17(3): 367-76.
von Kriegstein K, Smith DR, Patterson RD, Ives DT, Griffiths TD. Neural representation of auditory size in the human voice and in sounds from other resonant sources. CurrBiol 2007; 17(13): 1123-8.
von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD. Processing the acoustic effect of size in speech sounds. Neuroimage 2006; 32(1): 368-75.
Warren J, Scott S, Price C, Griffiths T. Human brain mechanisms for the early analysis of voices. Neuroimage 2006; 31(3): 1389-97.
Watson R, Latinus M, Charest I, Crabbe F, Belin P. People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. Cortex 2014; 50: 125-36.
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(5): 445-51.
Zimmermann P, Fimm B. Testbatterie zur Aufmerksamkeitsprüfung: TAP, Version 2.2: Psytest; 2009.

13.2 Supplementary Material

Supplementary Table 1. Individual details on each patient.

Patient	Age	Sex	MSO	Education (yrs)	Scan	Lesion type	Lesion site	Lesion size (*n* of voxels)	Aphasia	Cognitive disorder
1	66	f	157	10	MRT	TBI	l	43839	no	mild
2	36	f	151	12	MRT	TBI	r	9720	no	no
3	68	f	157	8	MRT	ICB	r	17606	no	no
4	40	m	273	10	MRT	TBI	r	5871	no	mild
5	63	m	148	10	MRT	SAH	r	2762	no	no
6	38	f	123	12	MRT	Tumour	l	39866	no	no
7	60	m	114	10	MRT	Isch stroke	l	20585385	no	no
8	63	m	80	10	MRT	Isch stroke	r	133889	no	no
9	27	f	84	12	MRT	TBI	r	15365	no	no
10	59	m	78	10	MRT	Isch stroke	l	115572	no	no
11	56	m	68	12	MRT	Isch stroke	l	84971	no	no
12	59	m	43	10	MRT	Isch stroke	r	45890	no	moderat
13	29	f	30	12	MRT	Isch stroke	l	39139	yes	mild
14	50	f		12	MRT	ICB	r	8612	no	mild
15	62	m	57	12	MRT	ICB	l	5122	no	no
16	22	f	54	10	MRT	ICB	r	73534	no	moderat
17	52	m	47	10	MRT	Isch stroke	l	8568	yes	mild
18	50	m	52	12	MRT	ICB	r	13150	no	moderat
19	43	m	48	10	MRT	Isch stroke	r	394740	no	mild
20	58	m	42	12	MRT	Isch stroke	r	17429	no	mild
21	49	m	109	8	MRT	TBI	r	23911	no	no
22	45	f	39	8	MRT	Isch stroke	r	4288	no	mild
23	46	f	40	10	MRT	SAH	r	663	no	mild
24	56	m	28	10	MRT	Isch stroke	r	277925	no	mild
25	61	f	33	10	MRT	Isch stroke	r	2930	no	mild
26	54	f	30	10	MRT	Isch stroke	l	50053	yes	no
27	31	f	31	10	MRT	Tumour	l	1367	no	mild

28	63	m	31	12	MRT	Isch stroke	r	853	no	mild
29	35	m	31	10	MRT	Isch stroke	r	185803	no	mild
30	56	f	24	10	MRT	Haem stroke	r	48246	no	moderat
31	53	f	25	12	MRT	Isch stroke	l	17748	yes	no
32	49	m	12	10	MRT	Isch stroke	r	25582	no	mild
33	50	f	35	10	MRT	Isch stroke	r	1413840	no	no
34	29	m	24	12	MRT	SAH	l	39718035	yes	mild
35	23	m	14	10	MRT	Isch stroke	r	53287	no	mild
36	57	f	18	10	MRT	Isch stroke	l	2203	no	mild
37	51	f	20	10	MRT	Aneurysma	r	6402	no	mild
38	31	f	17	10	CT	SAH	r	45674	no	mild
39	57	f	11	10	MRT	Isch stroke	l	183	no	mild
40	44	f	2	10	MRT	Isch stroke	l	11345	no	mild
41	49	f	10	12	MRT	Isch stroke	l	27363	yes	moderat
42	58	m	9	12	MRT	Isch stroke	l	27345	yes	moderat
43	29	m	3	10	MRT	Isch stroke	l	47795	yes	mild
44	55	f	5	10	MRT	Tumour	l	9433	no	mild
45	50	m	14	8	MRT	SAH	r	16879	no	no
46	48	m	11	10	MRT	Isch stroke	l	9625	no	mild
47	32	f	12	12	MRT	Tumour	r	611	no	mild
48	47	m	7	10	MRT	Isch stroke	r	32559	no	mild
49	58	f	10	12	CT	Isch stroke	r	200503	no	no
50	57	f	13	10	MRT	Isch stroke	r	6161	no	mild
51	41	f	25	8	MRT	Isch stroke	l	10670985	no	mild
52	39	m	7	10	MRT	Isch stroke	l	9573210	yes	mild
53	58	f	53	10	MRT	Isch stroke	l	224655	no	mild
54	37	f	9	12	MRT	TBI	l	15806	yes	no
55	42	m	23	9	MRT	Isch stroke	r	57669270	no	mild
56	51	f	9	12	MRT	SAH	l	76287	yes	mild
57	44	m	12	12	MRT	ICB	l	15893	yes	no
58	45	f	11	12	MRT	TBI	l	47892	yes	mild

MSO = months since onset, f = female, m = male, MRT = magnet resonance tomography, CT = computer tomography, TBI = Traumatic brain injury, SAH = Subarachnoid Hemorrhage, Haemorrhagic, ICB = Intracerebral Bleeding, isch stroke = ischaemic stroke, haem stroke = haemorrhagic stroke.

Familiar voice recognition: Famous voice test

Names of the famous Germans presented in the famous voice test (*n* = 42)

Marcel Reich-Ranicki	Otto Walkes	Til Schweiger
Joachim Löw	Helge Schneider	Anne Will
Günther Jauch	Michael Mittermeier	Alice Schwarzer
Ulrich Wickert	Franz Beckenbauer	Verona Pooth
Hellmuth Karasek	Boris Becker	Heidi Klum
Harald Schmidt	Michael Schumacher	Anke Engelke
Karl Lagerfeld	Gerhard Schröder	Barbara Schöneberger
Alfred Biolek	Helmuth Kohl	Sarah Kuttner
Thomas Gottschalk	Guido Westerwelle	Angela Merkel
Jürgen von der Lippe	Wolfgang Schäuble	Ursula von der Leyen
Oliver Pocher	Erich Honecker	Claudia Roth
Vicco von Buelow (Loriot)	Herbert Grönemeyer	Nena
Harpe Kerkelingen	Peter Maffay	Nina Hagen
Stefan Raab	Udo Lindenberg	Heike Makatsch

Acoustic voice processing: Vocal-pitch and vocal-timbre test

Stimuli and presentation software

The stimulus set consisted of five English vowels (/a/, /e/, /i/, /o/, /u/) resynthesised using the STRAIGHT software package (Kawahara *et al.*, 2008) implemented in Matlab (version 7.7, The MathWorks, Inc., MA, USA). The original vowels were spoken by a male speaker and had a duration of 600 ms (same material used in Smith *et al.* (2005)). For the pitch-discrimination task, all tokens of a given vowel were identical except for their fundamental frequency. F0 is the physical correlate of a speaker's glottal-pulse rate (GPR), which determines their voice pitch. For the timbre-discrimination task, all tokens of a given vowel were identical except for their spectral envelopes, which were scaled proportionally up or down in log-frequency space from the original spectral envelope. Spectral envelope is the physical correlate of a speaker's vocal-tract length (VTL), which is an aspect of vocal timbre that correlates with speaker size (Fitch and

Giedd, 1999). Both tests were implemented in Python (version 2.7.3, http://python.org/) and responses were recorded via a keyboard.

Procedure and analysis

We used an adaptive-tracking procedure (Kaernbach, 1991) to measure the participants' pitch and timbre just-noticeable differences (JNDs). During the pitch-discrimination test, participants listened to pairs of sequentially presented vowels differing only in their F0. One vowel per trial always had an F0 of 112 Hz, and the other was higher in F0 by an amount (ΔF0) defined in musical cents (1 semitone = 100 cents). The order of the stimuli was random on each trial, and participants reported which one was higher in pitch. The initial ΔF0 was 100 cents; this value decreased in steps of 10 cents following each correct response and increased in steps of 30 cents following each incorrect response. After four reversals (a switch from correct to incorrect response or vice versa within two consecutive trials), the up and down step sizes were changed to 6 and 2 cents, respectively, and the block of trials continued for a further 10 reversals. A JND was estimated for each single run by taking the mean of all ΔF0 values visited during the final 10 reversals. The participant's overall JND was defined as the mean JND over the five runs. Feedback about response accuracy followed each trial. At the beginning of the test, participants were familiarised with the auditory stimuli by presenting them with two vowels at the extremes of the F0 range. The average test duration was 15 minutes.

The timbre-discrimination test and analysis procedure was identical to the pitch-discrimination test except that the stimuli on each trial differed in their spectral envelopes, and participants reported which vowel was spoken by the smaller speaker. One vowel on each trial had a spectral envelope equal to that of the original speaker, while the other differed by ΔSER, defined in percent. Initial ΔSER was 12%; up and down step sizes were 3% and 1% for the first four reversals, and 0.6% and 0.2% for the remaining 10 reversals.

Face-recognition tests

Face-name test

Stimuli and presentation software

The stimulus set consisted of images of six male faces (three British and three Spanish actors who are not famous in Germany). The pictures were downloaded from freely accessible websites. The images were degraded by adding different levels of Gaussian noise to each pixel (e.g., 15%, 30%, 60%) using Adobe Photoshop CS4 (version 11.0.2; Adobe Systems, Inc., San Jose, CA, US). The complete stimulus set contains 228 images. Original photographs were taken from different views with varying facial expressions and under different lighting conditions. The test was implemented in the Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure and analysis

The test was structured into four learning and four interim testing phases, presented in alternating order, plus an additional final testing phase. During the learning phases, participants were presented with a sequence of face-name pairs (Peter, Jan, Timo, Alex, Otto, Leon). Participants had the task to learn the face-name associations. The learning phases differed in the number of presented face-name pairs: 10 face-name pairs per person during the first learning phase, three face-name pairs per person during the second and two face-name pairs per person during the third and fourth learning phase. During the interim-testing and final testing phases, the participants were presented with a novel face image of one the learned identities together with the six names. Participants were asked to select the corresponding name for the face. Each testing phase contained five items per identity. During the second, third and fourth interim-testing phases, the face pictures were presented with increasing Gaussian noise levels (15%, 30% and 60% respectively) to minimize ceiling effects. During the final testing phase, again no noise was added to the face images. In the first and second interim testing phase, participants received feedback on their decision and the correct face-name pair was presented again. The whole test took approximately 30 minutes.

As a measure of face-name recognition performance, we calculated the average percent correct over the five testing phases.

Supplementary results

Supplementary Table 2. Comparison of behavioural test results for patients with left and right hemispheric lesions.

Tests	*n*	RH patients	*n*	LH patients	Statistics	p value
Voice recognition (newly-learned and familiar voices) [z]	20	-0.11 (0.77)	22	0.06 (0.55)	0.96	0.339
Newly-learned voice recognition						
Newly-learned voice test [%]	23	57.33 (11.08)	20	63.00 (8.63)	-2.29	0.022**
Voice-name test [%]	31	54.44 (12.22)	27	58.06 (13.63)	-1.15	0.252
Voice-face test [%]	23	60.93 (12.53)	20	67.70 (9.63)	-2.20	0.027*
Familiar voice recognition						
Famous voice test [z]	30	0.095 (0.76)	27	-0.11 (0.69)	1.08	0.303
Familiarity judgment d’	30	1.22 (0.89)	27	0.98 (0.77)	-1.24	0.215
Semantic association [%]	30	70.57 (19.94)	27	68.46 (17.78)	-0.60	0.549
Acoustical voice tests						
Timbre test [SER]	30	10.40 (4.24)	25	8.92 (5.01	-1.00	0.319
Pitch test [cent]	31	116.16 (58.48)	27	135.52 (73.54)	1.25	0.269
Visual control tests						
CFMT [%]	31	64.11 (16.76)	26	69.98 (13.53)	2.06	0.157
Face-name test [%]	23	55.78 (15.84)	20	66.01 (21.47)	3.21	0.080

The table displays mean (standard deviation) values on each behavioural test for patients with left and right hemispheric lesions. Statistical group differences for tests with normal distributed values were computed by ANOVA (i.e., familiar voice recognition, pitch test, CFMT, face-name test) and for tests without normal distributed values by a Mann-Whitney test (i.e., unfamiliar voice recognition, voice-name test, voice-face test, famous voice test, timbre test). Statistics for ANOVA analyses refer to F values and for Mann-Whitney test analysis to Z values. ** indicates significant group differences at $p < 0.05$ after Bonferroni correction, *indicates significant group difference at $p < 0.05$ (uncorrected).

Correlation between neuropsychological and behavioural measures

Of the 99 correlations we run, 89 were non-significant ($p < 0.05$, uncorrected). In turn, 11 behavioural measures were significantly correlated at an uncorrected level with demographic or neuropsychological scores, i.e. voice-name test with age and hearing level; famous voice test (familiarity feeling, semantic association) with time since lesion onset and hearing level (only semantic association); vocal-pitch test with education, digit span, and German Vocabulary Test; CFMT with face-name learning (cued recall and memory score); and face-name test with time since lesion onset.

Self-assessment on every-day person-recognition abilities after lesion onset

Based on a Friedman test (at $p < 0.001$) abilities for voice, face, and general person recognition were worse after lesion onset as compared to the time before brain injury (Table S4). 13 patients reported reduced abilities in person recognition after lesion onset. 11 patients reported decreased voice-recognition and 14 patients decreased face-recognition abilities after lesion onset. We checked if self-assessments after lesion onset and person-recognition performance of the behavioural test battery were correlated. There was a correlation between self-assessment of person recognition abilities and the ability to judge voice familiarity ('familiarity decision' of the famous voice test). There was also a correlation between the self-assessment of voice recognition and the face-name test performance (at $p < 0.01$). Patients reported to rely on the same features to recognise a person before and after lesion onset (Table S5).

Supplementary Table 3. Self-assessment on person-recognition abilities in every-day situations.

	Person recognition*		**Voice recognition***		**Face recognition***	
	Pre (RBD/LBD)[1]	Post	Pre[1]	Post	Pre[1]	Post
Excellent	21 (10/11)	16 (8/8)	12 (5/7)	11 (3/8)	19 (10/9)	13 (4/9)
Good	20 (12/9)	25 (13/12)	24 (13/11)	24 (13/11)	21 (12/9)	30 (17/13)
Sufficient	1 (1/0)	12 (6/6)	8 (5/3)	18 (11/7)	4 1/3)	12 (8/4)
Poor	2 (0/2)	5 (4/1)	0	5 (4/1)	0	3 (2/1)
Not at all	0	0	0	0	0	0

*Significantly worse self-assessment of person-, voice-, and face-recognition abilities before and after lesion onset (at $p < 0.001$, Friedman Test). ^ RBD indicated significantly worse voice recognition abilities after lesion onset compared to LBD patients (One-way-anova, $p = 0.37$)).

[1] For the pre-assessment, 14 cases are missing because they did an older version of the questionnaire where only post-abilities were assessed.

Supplementary Table 4. Self-assessment on features used for person recognition.

Person-recognition features	**Pre[1]**	**Post**
Voice	0	2
Face	13	15
Voice and face	25	29
others	6	12

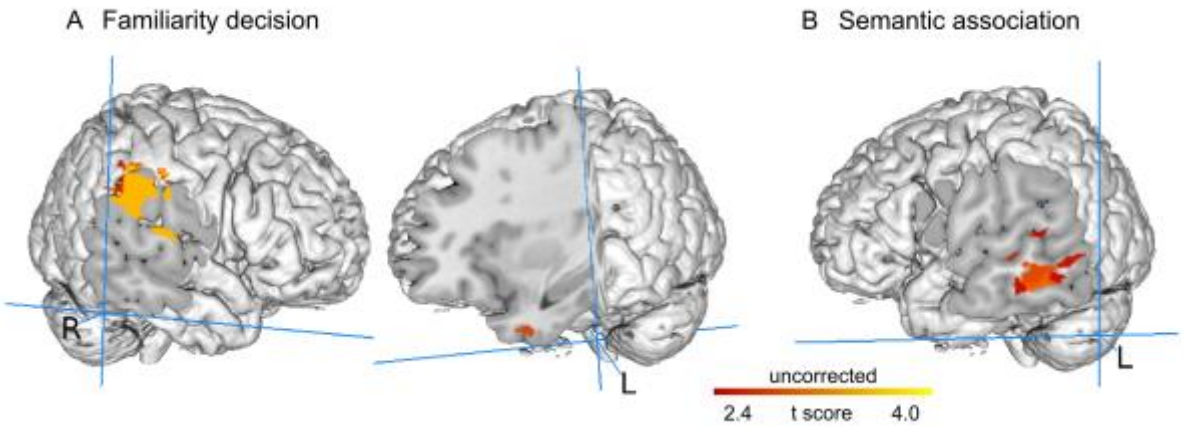
Features used for person recognition did not change significantly before and after lesion onset (at $p < 0.05$, Friedman Test).

[1] For the pre-assessment, 14 cases are missing because they did an older version of the questionnaire where only post-abilities were assessed.

Supplementary Table 5. Overview of brain structures lesioned in more than 3 patients within the ROI maps.

	Covered	**Not-covered**
Bilateral temporal lobe map		
With a-priori anatomical hypothesis	Bilateral STG, bilateral MTG, bilateral Heschl's gyrus, bilateral planum temporale	
Without a-priori anatomical hypothesis		Very anterior/ medial right temporal pole, bilateral visual areas such as anterior inferior temporal gyrus, posterior temporal fusiform cortex, and temporal occipital fusiform cortex
Right inferior parietal lobe map		
With a-priori anatomical hypothesis	Angular gyrus, supramarginal gyrus	
Without a-priori anatomical hypothesis		Superior lateral occipital cortex and the postcentral gyrus

Supplementary Figure 1. Results of the sub-scores of the famous voice test.



A. With familiarity decision associated lesion clusters were located in the right inferior parietal lobe (i.e. anterior SMG) and in left temporal lobe (i.e. left temporal). B. For semantic association, we observed a brain-behaviour association in two lesion clusters in the left temporal lobe (i.e. posterior MTG, posterior STG). Voxels shown exceeded a more lenient threshold for significance ($p < 0.01$ uncorrected, voxel size > 100).

Supplementary Table 6 Overview of lesion-cluster coordinates associated with the sub-scores of the famous voice test.

	Familiarity decision						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Right parietal lobe												
anterior SMG	64	-26	28	3.52	-	2370	-	-	-	-	-	-
Left temporal lobe												
Temporal pole	-32	5	-37	2.83	-	154	-	-	-	-	-	-
	Semantic association						+ HL and LV					
	x	y	z	Max T	*p*	*n* of voxels	x	y	z	Max T	*p*	*n* of voxels
Left temporal lobe												
posterior MTG	-55	-36	-10	3.35	-	6067	-	-	-	-	-	-
posterior STG	-63	-37	10	2.58	-	123	-	-	-	-	-	-

Results of the familiarity decision and semantic association sub-score are reported at $p < 0.01$ uncorrected, voxel size > 100. SMG = supramarginal gyrus, STG = superior temporal gyrus. MTG = middle temporal gyrus. HL = hearing level, LV = lesion volume.

Supplementary references

Fitch WT, Giedd J. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am 1999; 106(3 Pt 1): 1511-22.

Kaernbach C. Simple adaptive testing with the weighted up-down method. Perception & psychophysics 1991; 49(3): 227-9.

Kawahara H, Morise M, Takahashi T, Nisimura R, Irino T, Banno H. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. Acoustics, Speech and Signal Processing, 2008 ICASSP 2008 IEEE International Conference on; 2008 March 31 2008-April 4 2008; 2008. p. 3933-6.

Smith DRR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sounds. JAcoustSocAm 2005; 117(1): 305-18.

14 DANKSAGUNG

Ich möchte allen voran Claudia Roswandowitz danken für ihre ausdauernde Betreuung, Unterstützung und vielen Denkanstöße sowie die prägende gemeinsame Zeit während der Studie bis heute. Großer Dank gilt außerdem Hellmuth Obrig und Katharina von Kriegstein für Betreuung, Hilfe bei der Durchführung der Studie und Antworten auf viele Fragen.

Weiterhin möchte ich den KollegInnen der Arbeitsgruppe danken – insbesondere Stefanie Schelinski, Kamila Borowiak und Sonja Schall – für viele gemeinsame Momente, lebhafte Gespräche, Anregungen und Ermunterungen.

Ich danke Tobias Albert für seine Hilfe bei der Grafik und das Da-Sein als mentale Stütze. Meinen Eltern danke ich für finanzielle und seelische Unterstützung, unermüdliche Nachfragen und den festen Glauben an mich. Ein weiterer Dank gilt Hannah Kappes, ohne die vermutlich keine Nummerierung stimmen würde.