 doi:10.20944/preprints201806.0280.v2

BOOKCHAPTER

Title

Deficits in Voice-Identity Processing: Acquired and Developmental Phonagnosia

Author names

Claudia Roswandowitz [a,b] *

Corrina Maguinness [a] *

Katharina von Kriegstein [a,c,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] Humboldt University zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

[d] Technische Universität Dresden, Faculty of Psychology, Bamberger Str. 7, 01187 Dresden

*both authors contributed equally to this work

Correspondence should be addressed to Claudia Roswandowitz (roswandowitz@cbs.mpg.de) and Katharina von Kriegstein (katharina.von_kriegstein@tu-dresden.de)

This is a draft of a chapter that has been accepted for publication by Oxford University Press in the forthcoming book The Oxford Handbook of Voice Perception edited by Sascha Frühholz and Pascal Belin due for publication in November 2018.

 doi:10.20944/preprints201806.0280.v2

Abstract

The voice contains elementary social communication cues, conveying speech, as well as paralinguistic information pertaining to the emotional state and the identity of the speaker. In contrast to vocal-speech and vocal-emotion processing, voice-identity processing has been less explored. This seems surprising, given the day-to-day significance of person recognition by voice. A valuable approach to unravel how voice-identity processing is accomplished is to investigate people who have a selective deficit in recognising voices. Such a deficit has been termed *phonagnosia*. In the present chapter, we provide a systematic overview of studies on phonagnosia and how they relate to current neurocognitive models of person recognition. We review studies that have characterised people who suffer from phonagnosia following brain damage (i.e. acquired phonagnosia) and also studies, which have examined phonagnosia cases without apparent brain lesion (i.e. developmental phonagnosia). Based on the reviewed literature, we emphasise the need for a careful behavioural characterisation of phonagnosia cases by taking into consideration the multistage nature of voice-identity processing and the resulting behavioural phonagnosia subtypes.

Keywords: phonagnosia, acquired, developmental, apperceptive, associative, voice-identity processing, speaker recognition, core-voice system, extended system

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Introduction

Recognising a person by voice is a skill, which humans master with ease. However, for some people this skill can be impaired. This deficit is termed 'phonagnosia' (Van Lancker and Canter, 1982); originating from the Greek words 'φώνημα' or 'phone' meaning voice or sound and the term agnosia (αγνώσις). Agnosia is commonly used for conditions in which the recognition of stimuli is disturbed (Lissauer, 1890; Freud, 1891). In phonagnosia, the ability to process other vocal information e.g., gender, age, and emotion as well as speech, music, and facial information is largely preserved (Neuner and Schweinberger, 2000; Garrido *et al.*, 2009; Roswandowitz *et al.*, 2014). Phonagnosia can occur after brain damage (i.e. *acquired phonagnosia*) (Assal *et al.*, 1976; Van Lancker and Canter, 1982; Neuner and Schweinberger, 2000) or in the absence of brain insult (i.e. *developmental phonagnosia*) (Garrido *et al.*, 2009; Roswandowitz *et al.*, 2014). The disorder has currently two major sub-classifications: apperceptive and associative phonagnosia. In apperceptive phonagnosia, the deficit lies in the perceptual analysis of voice features, whereas the association of semantic information to a voice is intact (Hailstone *et al.*, 2011; Roswandowitz *et al.*, 2014; Xu *et al.*, 2015). While associative phonagnosia is understood as a failure to recognise a voice as familiar (familiarity decision) and to associate semantic information to a voice (semantic processing), though the perception of the voice is unaffected (Hailstone *et al.*, 2010; Hailstone *et al.*, 2011; Roswandowitz *et al.*, 2014).

Though phonagnosia may offer a unique instance to study auditory person recognition, the number of scientific investigations so far has been limited. This might be caused by the following factors: (i) Phonagnosia has been under scientific investigation for a rather short time. The first study on acquired phonagnosia was published in 1976 (Assal *et al.*, 1976) and on developmental phonagnosia in 2009 (Garrido *et al.*, 2009). (ii) Testing of voice-recognition deficits is relatively difficult, as standard tests are not readily available and are often language-dependent (but see Aglieri *et al.*, 2017). (iii) Cases of phonagnosia are rare, although this perceived rarity may be more related to a low self-awareness, rather than a low prevalence rate, of voice-identity processing disorders (Roswandowitz *et al.*, 2014).

In the following chapter, we provide a systematic overview of investigations on phonagnosia and how they relate to current models of voice-identity processing. We begin by introducing a neurocognitive model of voice-identity processing and provide an overview of the

 doi:10.20944/preprints201806.0280.v2

behavioural tests, which are used to assess cases and subtypes of phonagnosia. We then review clinical studies, which documented cases of acquired phonagnosia, before turning to focus on recently reported cases of developmental phonagnosia. We discuss the reviewed findings within the context of current voice-identity processing models and conclude with proposing future research directions.

Model of voice-identity processing

Recognising voices at the individual level is a challenge for the perceptual and cognitive system. Each voice that we hear shares the same basic perceptual features across individuals (acoustic parameters such as pitch and timbre (Lavner *et al.*, 2001; López *et al.*, 2013)); and thus the brain is tasked with representing a unique voice in memory, by perceiving and representing often subtle differences in these features across individuals (Belin *et al.*, 2011). Furthermore, it is not sufficient that we simply recognise a voice as familiar. Rather, successful voice recognition also involves linking the familiar voice to stored knowledge, or semantics, including where we know the voice from, what the person looks like, are they a friend or a foe? Thus, voice recognition can be conceived as a *multistage* process, which begins with the encoding of the incoming vocal signal and ends in successful identification of the voice at the level of a specific individual identity. In Figure 1A we present a cognitive model of voice-identity processing and highlight candidate brain regions in Figure 1B, which may support this multistage process. We also outline how subtypes of phonagnosia, apperceptive and associative, may arise due to dysfunction at different *stages* of voice-identity processing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

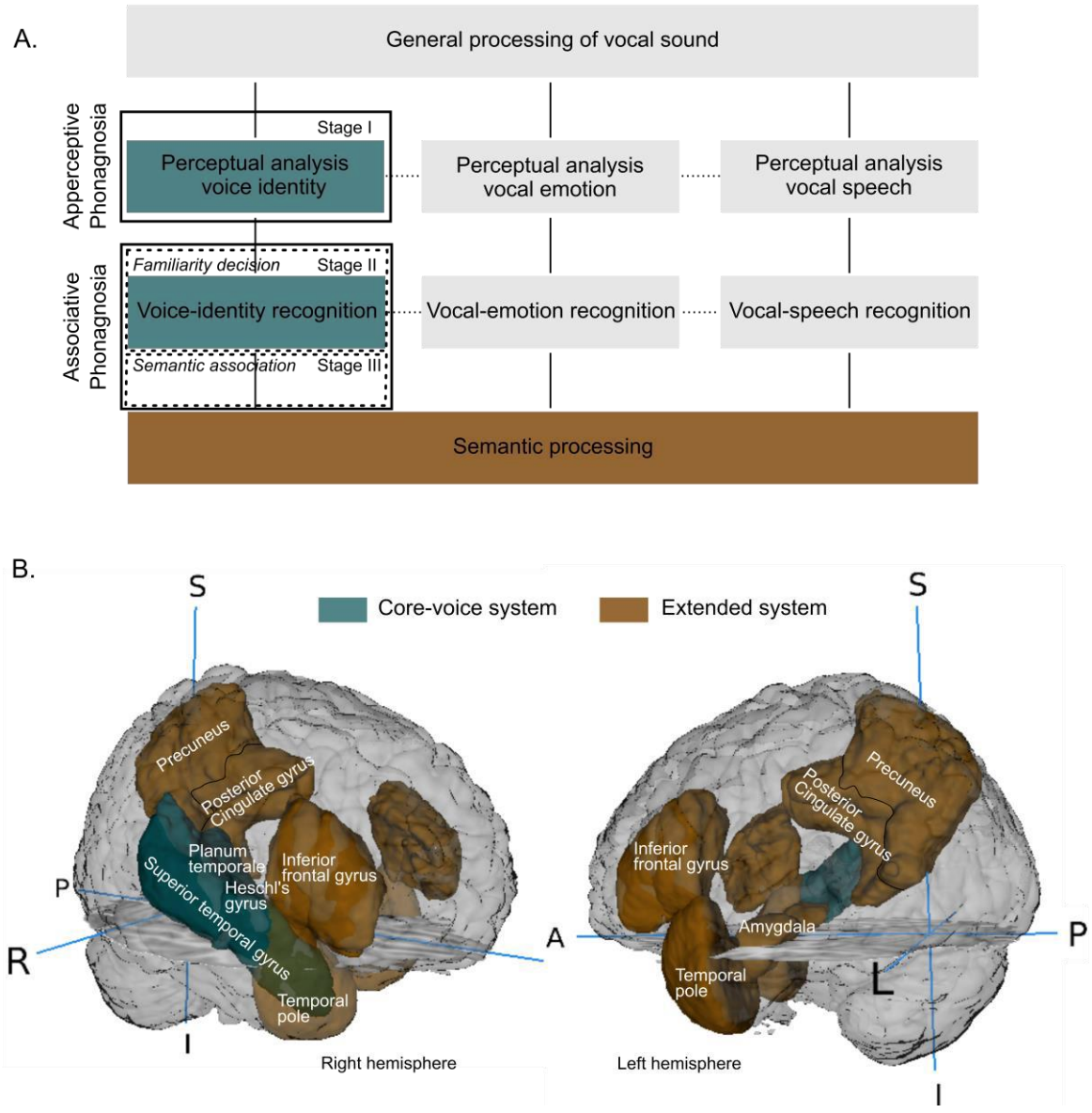

Figure 1. Neurocognitive model of voice processing. A. A model (adapted from Ellis *et al.* (1997), Belin *et al.* (2004), Blank *et al.* (2014), and Neuner and Schweinberger (2000), based on a seminal model of face processing outlined by Bruce & Young, (1986)) which describes the cognitive processes involved in voice-identity processing. B. Overview of potential brain structures supporting voice-identity processing, as evidenced in neuroimaging studies with neurotypical participants. R= Right; L = Left; S = Superior; I = Inferior; A = Anterior; P = Posterior.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

According to the model (Figure 1A), the vocal sound undergoes an initial general processing phase. This processing may be partly shared and partly independent from the processing of other sound sources, including object sounds or music. After this initial phase, voice-identity processing begins. Stage I: Here, the perceptual system analyses complex spectrotemporal acoustical properties of the incoming vocal sound, which support identity processing. This stage encompasses 'structural encoding' (see e.g. Neuner and Schweinberger, 2000), where invariant properties of the voice (vocal properties which remain constant across different speech utterances or changes in prosody) are extracted. These properties are merged to create a coherent voice percept. The merged voice properties may be contrasted against a 'prototype' voice (Lavner *et al.*, 2001; Andics *et al.*, 2010; see prototype encoding of voices; Latinus *et al.*, 2013; for review see Maguinness *et al.*, 2018). The prototype voice may represent an average approximation of the voices the listener has encountered or it may reflect a "very common voice" (Lavner *et al.*, 2001). The computed acoustical differences between the voice percept and the prototype voice can be passed on for analysis to support identity- recognition at later stages of processing. Other features of the vocal sound, which support vocal emotion and speech processing are also analysed at this stage but are argued to be processed in partly independent but interacting systems (von Kriegstein *et al.*, 2010; Kreitewolf *et al.*, 2014). The stage I of processing is suggested to be supported by brain regions of a core-voice system. Potential candidate brain regions are the posterior and mid regions of superior temporal gyrus/sulcus (STG/S) (e.g., Belin *et al.*, 2000; von Kriegstein and Giraud, 2004; Warren *et al.*, 2006; Pernet *et al.*, 2015; Roswandowitz *et al.*, 2018) and auditory regions such as the planum temporale (von Kriegstein *et al.*, 2006b; Warren *et al.*, 2006) and Heschl's gyrus (Formisano *et al.*, 2008; Bonte *et al.*, 2014), predominantly in the right hemisphere (Figure 1B). Apperceptive phonagnosia may emerge due to dysfunction at this early stage of processing (Figure 1A, Stage I). Poor perceptual analysis of the voice may result in a weak representation of the voice- individuating features, which may impact negatively on later stages of processing (voice-identity recognition).

Stage II: At the stage of voice-identity recognition, a sense of familiarity is generated if the computed voice percept closely resembles a stored voice representation. These voice representations may be stored as relatively unique 'reference patterns' for each known voice-identity (see Lavner at al. 2001). This process is likely supported by anterior and mid regions of

 doi:10.20944/preprints201806.0280.v2

STG/S including parts of the anterior temporal lobe (most likely superior lateral part, see e.g. Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003) in the core-voice system, while more posterior regions are concerned with perceptual voice analysis (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; Andics *et al.*, 2013; Latinus *et al.*, 2013; Schall *et al.*, 2015). A deficit at this stage would give rise to deficient familiarity decisions despite a successfully analysed vocal percept. We will call this familiarity-associative phonagnosia (Figure 1A, Stage II). Disrupted access to the stored voice-identity representations constrains the ability to judge whether the voice has been encountered before.

Stage III: After the voice has been recognised as familiar it is linked to stored multi-modal semantic information characterising the person identity. This multi-modal information is processed in an extended system (semantic processing), which is proposed to share connections with the core-voice system. Regions concerned with vocal emotion and speech recognition may also share connections with this extended system. Potential brain candidates for the extended system include supra-modal regions encompassing discrete regions of the temporal pole, precuneus/posterior cingulate, amygdala, and inferior frontal gyrus (Shah *et al.*, 2001; von Kriegstein and Giraud, 2006; Andics *et al.*, 2010; Latinus *et al.*, 2011; for review see Blank *et al.*, 2014). Dysfunction at this stage of processing, i.e. poor connectivity between the core-voice and extended system (Figure 1A, Stage III), may underpin cases of semantic-associative phonagnosia which are characterised by a deficit in associating semantic information to a voice, which has been successfully perceived and categorised as familiar[1]. Note that we focus here on the auditory modality, for reviews on how voice information is linked to face representations at several stages of processing see (von Kriegstein, 2011; Blank *et al.*, 2014; Maguinness and von Kriegstein, 2017).

[1] The classification of subtypes of phonagnosia is informed by the visual agnosia literature (Lissauer, 1890; De Renzi *et al.*, 1991). There, an 'apperceptive' agnosia is consistently categorised as a perceptual processing deficit (i.e. Figure 1A, Stage I) (Warrington, 1975; De Renzi, 1986; De Renzi *et al.*, 1991). However, there is much discrepancy regarding the definition of 'associative' agnosia, specifically within the realm of prosopagnosia ('face-blindness'), a visual parallel disorder to phonagnosia. Classically, associative (prosop)agnosia has been defined as a failure to link an analysed percept to stored multi-modal semantic information (i.e. Figure 1A, Stage III) (Warrington, 1975; Warrington and Shallice, 1984). However, others have stated that this poor semantic association should be labelled 'amnestic' prosopagnosia and that 'associative' prosopagnosia rather reflects a failure to link the analysed percept to a stored facial representation i.e. impaired familiarity decisions (i.e. Figure 1A, Stage II) (Fox *et al.*, 2008;

 doi:10.20944/preprints201806.0280.v2

Stollhoff *et al.*, 2011; Avidan and Behrmann, 2014). Here, we propose to resolve the discrepancy by adopting the general label of 'associative phonagnosia' as a deficit, which encompasses a failure to attribute meaning to the successfully analysed vocal percept. This may arise due to 1) impaired familiarity decisions or 2) impaired semantic association to the vocal identity. To avoid confusion, we will call the first familiarity-associative phonagnosia and the second semantic-associative phonagnosia.

Tests of voice-identity processing

Given the theoretical framework proposed in Figure 1A, tests for phonagnosia need to be designed to address the multistage nature of voice-identity processing. Currently, employed voice-processing tests (summarised in Table 1) include measures which can evaluate: 1) the perceptual analysis of the vocal signal, achieved through means of unfamiliar voice discrimination and unfamiliar speaker change detection tests. Such tests can reveal apperceptive impairments; 2) a sense of familiarity with the encoded familiar vocal percept (i.e. familiarity decision); 3) the ability to link the encoded familiar vocal percept to identity-specific person knowledge (i.e. semantic association). Familiar voice-recognition tests are commonly used to examine both familiarity decisions and semantic association. These tests often assess both associative abilities, e.g. first listeners indicate a sense of familiarity towards a voice and then associate semantic knowledge to the familiar voice. Voices presented in those tests may involve famous, personally familiar or newly learned speakers' voices.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Table 1. Overview of the tasks and stimuli used for assessing apperceptive (upper section) and associative (lower section) voice-identity processing.

Design	**Stimuli** *Voice familiarity*	*Test stimuli*	**Paradigm**	**Reference**
Apperceptive voice-identity processing				
Voice discrimination	Unfamiliar voices	3-word sentences Test 1: 2 male, 2 female, 2 child speakers Test 2: 5 females speaking English, German, Spanish, Italian, Japanese Test 3: 5 young female French speakers	Presentation of 30 pairs of sentences, Same-different speaker judgment, Pairs of sentences had either same or different word content	Assal et al., 1976
	Unfamiliar voices	Sentences Test 1: Male, female, child speakers Test 2: French female speakers Test 3: Hebrew female speakers	Presentation of 40 pairs of sentences, Same-different speaker judgment	Assal et al., 1981
	Unfamiliar voices	Sentences 10 male speakers	Presentation of 26 pairs of sentences, Same-different speaker judgment, Pairs of sentences same content, if same speaker presented different tokens used	Van Lancker et al., 1987, 1988, 1989
	Unfamiliar voices	Sentences (2 sec long) Unfamiliar speakers	Presentation of 54 pairs of sentences, Same-different speaker judgment, Pairs of sentences different content, same gender	Neuner and Schweinberger, 2000
	Unfamiliar voices	Sentences with 3 key words 6 female speakers	Presentation of sentence pairs in 4 different SNRs: -6, 0, 6, or 12 dB (24 sentence pairs per SNR), Same-different speaker judgment	Garrido et al., 2009

 doi:10.20944/preprints201806.0280.v2

	Unfamiliar voices	Sentences 21 female speakers	Presentation of NV sentence pairs with 6, 16, or 48 frequency channels (24 sentence pairs per frequency channel level) Same-different speaker judgment	Garrido et al., 2009
	Unfamiliar voices	2-word sentences 3 male speakers	Brief familiarisation with the voice identities (passive listening), followed by: Presentation of 54 pairs of sentences, Same-different speaker judgment	Roswandowitz et al., 2014
	Unfamiliar voices	Sentences 5 female speakers	Presentation of target voice, followed by: (i) 5, 10 or 20 second interval, and (ii) presentation of 2 test voices (40 trials per interval duration) 2AFC speaker matching task	Xu et al., 2015
Speaker change detection	Unfamiliar voices	4 min long text	Text included 24 speaker changes, Speaker change detection	Assal et al., 1981
	Unfamiliar voices	High frequent words (names of weekdays and months) Female speakers	Sequences of words including speaker changes - 24 trials of weekdays - 24 trials of months, Speaker change detection	Hailstone et al., 2010
	Unfamiliar voices	High frequent words (names of weekdays) Female speakers Test 1: Naturalistic stimuli Test 2: Fixed f0 (220Hz)	Sequences of words including speaker changes Test 1: 28 trials Test 2: 12 trials, Speaker change detection	Hailstone et al., 2011
Associative voice-identity processing				
Familiar voice recognition	Famous voices	7 celebrity voices, Male speaker	For each voice, cross-modal matching on 4-choice response array (voice – face/name) (*semantic association*)	Van Lancker et al., 1982

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Famous voices	25 celebrity voices, Male speaker, 4 sec long samples	For each voice, cross-modal matching on 4-choice response array (voice – face/name) (*semantic association*), Debriefing of subjective familiarity with celebrities (van Lancker et al., 1989)	Van Lancker et al., 1987, 1988, 1989
Famous and unfamiliar voices	32 celebrity voices, 32 unfamiliar voices, Female and male speaker, 2 sec long samples	After voice presentation, (i) familiarity judgment (*familiarity decision*), (ii) if familiar, voice naming (*semantic association*)	Neuner and Schweinberger 2000
Personally familiar and unfamiliar voices	Per participant: 1 familiar voice, 5 unfamiliar voices, Voice samples consisted of vowels, CVC syllables, words, and sentences	After voice presentation, familiarity judgment (*familiarity decision*)	Lang et al., 2009
Famous and unfamiliar voices	48 celebrity voices 48 unfamiliar voices 7 sec long samples	After voice presentation, (i) familiarity judgment (*familiarity decision*), (ii) if familiar, voice identification (provide name or other biographical detail) (*semantic association*)	Garrido et al., 2009
Famous and unfamiliar voices Famous and unfamiliar voices	24 celebrity voices, Female and male voices 24 celebrity voices (same as above)	After voice presentation, familiarity judgment (*familiarity decision*) After voice presentation, (i) voice identification (provide name or other biographical detail) (ii) cross-modal matching on a response array (voice – face/name) (*semantic association*)	Hailstone et al., 2010, 2011 Hailstone et al., 2010, 2011
Famous and unfamiliar voices	42 celebrity voices, 20 unfamiliar voices, Female and male voices, 5 sec long samples	After voice presentation (i) familiarity judgment (*familiarity decision*) (ii) if familiar, voice identification (provide name or other biographical detail; Roswandowitz et al., 2014)	Roswandowitz et al., 2014, 2018

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

			cross-modal matching on a response array (voice – face/name; Roswandowitz et al. 2018) (*semantic association*), Debriefing of subjective voice familiarity of celebrities	
	Famous and unfamiliar voices	100 celebrity voices 100 unfamiliar voices 6-8 sec long samples	Celebrity face-name composite displayed (1,2 or 4 identity composites), followed by 2 voice samples, then: (i) indicate which of the voices is a celebrity (*familiarity decision*) (ii) cross-modal matching of famous voice to face/name composite (1,2 or 4 options) (*semantic association*)	Xu et al., 2014, see also Herald et al., 2014
Newly-learned voice recognition	Newly-learned voices	6 unfamiliar female speakers Sentences with 3 key words	Cross-modal learning: name and voice, followed by: (i) Voice recognition task (is the voice the same as the target speaker?) (ii) Voice recognition task (what name matches the voice identity? 6 options) (iii) Old/new task (is the voice new or has it been heard before i.e. old?)	Garrido et al., 2009
	Newly-learned voices	3 male and 3 female unfamiliar speakers per test (voice-name/ voice-face test)	Voice-name test: Cross-modal learning: simultaneous voice-name presentation, Testing: voice - name matching Voice-face test: same paradigm, just with voice-face associations	Roswandowitz et al., 2014, 2018

SNR = signal-to-noise ratio, NV = noise-vocoded, AFC = alternative forced choice

 doi:10.20944/preprints201806.0280.v2

Acquired phonagnosia

Lesion studies on phonagnosia allow a strong interpretation about brain regions required to identify voices. In the following, we review brain lesion studies which aimed to characterise the cognitive and neural mechanisms supporting voice-identity processing. The term phonagnosia implies a modality-specific deficit requiring many different control tests. However, the number of control tests or self-reports assessing other person-recognition or speech-processing abilities varies across clinical studies. Thus, whether the reported cases of acquired phonagnosia are also associated with other impairments often remains unclear in particular in those studies that do not include a systematic investigation of control abilities. For an overview of the reviewed lesion studies on voice-identity processing see Table 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Table 2. Overview about lesion studies on voice-identity processing.

Study	**Subjects**				**Lesion**		**Brain imaging**	**Behavioural findings**			**Brain-behaviour findings**
	n	*Age*	*MSO*	*Hearing*	*Type*	*Location*		*Voice*	*Face*	*Speech*	
Assal et al., 1976	47 patients 22 LBD 25 RBD 29 controls	 45 46 43	N/A	Normal hearing (self-reports)	Vascular diseases (12), tumor/abscess (7), TBI (3)	 LH RH	N/A	RBD: impaired unfamiliar voice DISCR Acquired voice DISCR deficit exist	N/A	LBD: aphasia no affect on voice DISCR (=independent)	RH = voice DISCR deficit
Assal et al., 1981	76 patients 40 LBD 36 RBD 35 controls	N/A	N/A	N/A	N/A	 LH RH	N/A	RBD: impaired unfamiliar voice DISCR	N/A	LBD: aphasia no affect on voice DISCR (=independent)	RH = voice DISCR deficit
	52 patients 28 LBD 24 RBD 11 controls	N/A	N/A	N/A	N/A	 LH RH	N/A	RBD: impaired speaker change detection	N/A	N/A	RH = speaker change detection deficit
	Case RB	45	N/A	Normal hearing	Vascular diseases	Bilateral anterior TL, settling after 6 months in left TL	N/A	Impaired familiar voice REC, Moderately impaired unfamiliar voice DISCR, (Amusic)	Intact (selfreport)	N/A	TL = voice REC deficit
Van Lancker et al., 1982	30 patients 21 LBD 9 RBD	 62 52	 8.92	Hearing sufficient for speech perception	Cerebral vascular diseases, TBI	 LH RH	Neurological evaluation, CAT scans	RBD: impaired familiar voice REC	RBD: impaired familiar face REC	All LBD aphasic	RH = voice and face REC deficit = Voice and face REC tend to co-occur (termed phonagnosia)
Van Lancker et al., 1987	32 patients 15 LBD 11 RBD	 61 59	 2-24 2-12	Normal hearing	Stroke (40), Craniotomies (2),	 LH RH	CAT scans, EEGs,	RBD: impaired familiar voice REC and	Intact face recognition in 4 patients	All LBD aphasic	RH = voice REC deficit,

	6 BBD 48 controls	69 64	1-24		haemorrhage (1), meningioma (1), tumor (1)	Both Hs	neurological evaluations	unfamiliar voice DISCR, LBD: impaired unfamiliar voice DISCR, intact familiar voice REC in 14 patients dissociation between familiar voice REC and unfamiliar voice DISCR			LH + RH = voice DISCR deficit
Van Lancker et al., 1988	6 case reports 2 LBD 1 RBD 3 BBD 30 controls	65 (52 – 82) 50-85	0.5 – years after TSO	Normal hearing reported for 2 cases	Stroke (5) Haemorrhage (1)	Mainly temporal, parietal, frontal lobe	CTs	4 BBD: impaired unfamiliar voice DISCR, 3 RBD: impaired familiar voice REC, 1 RBD: impaired in both tasks In 5 patients dissociation between familiar voice REC and unfamiliar voice DISCR	N/A	4 LBD with aphasia and voice DISCR deficit	Right PL = voice REC deficit, Bilateral TL = voice DISCR deficit
Van Lancker et al., 1989	56 patients 25 LBD 25 RBD 6 BBD 48 controls	 61 63 71 64	N/A.	N/A	Cerebral infarction	Lesions classified in parietal, temporal, and temporo-parietal lesions	CTs	BBD: impaired unfamiliar voice DISCR, RBD: impaired familiar voice REC	N/A	All LBD aphasic	Quantitative evidence for: Right PL = voice REC deficit, Bilateral TL = voice DISCR deficit

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Neuner and Schweinberger, 2000	36 patients 16 LBD 13 RBD 7 BBD 20 controls	48 44	8.2	N/A	anemic infarct (10), haemorrhage (10), subarachnoid haemorrhage (5), TBI (5), hypoxia (2), TBI with hypoxia (1), encephalitis (2), tumor (1)	LH RH Both Hs	Surgery reports, CTs, or MRI scans	4 patients: selective voice REC deficit (intact face, name, and sound REC), 2 RBD, 1 LBD: impaired in familiar voice REC, 1 RBD: impaired in familiar voice REC and unfamiliar voice DISCR	In 4 patients intact face-identity processing	N/A	RH = voice REC deficit
Lang et al., 2009	20 patients 11 LBD 9 RBD 17 controls	 66 64 64	 3.1 1.6	N/A	Ischemic infarcts	LH: MCA (6), PCA (2), LSA (3) RH: MCA (9), PCA (1)	N/A	RBD: impaired familiar voice REC LBD: intact familiar voice REC	N/A	All LH aphasic and intact voice REC	RH = voice REC deficit
Hailstone et al., 2010	Case QR	61	-	Normal hearing	Behavioural variant frontotemporal dementia	Right anterior TL extending to TL (STG)	MRI scans	Impaired familiarity and REC of familiar voices, Intact unfamiliar voice DISCR (impaired music instrument processing)	Impaired familiarity, moderately impaired REC of familiar faces, Intact unfamiliar face DISCR	N/A	Right anterior TL and STG = voice REC deficit (associative phonagnosia)
	Case KL 24 controls	72	-	Normal hearing	Frontotemporal lobar degeneration	Bilateral anterior TL atrophy extending to inferior temporal cortices (incl. FFA)	MRI scans	Impaired familiarity and REC of familiar voices, Intact unfamiliar voice DISCR	Impaired familiarity and REC of familiar faces, Intact unfamiliar face DISCR	N/A	Bilateral anterior TL = multi-modal person REC deficit (voice, face, name)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Hailstone et al., 2011	36 patients 14 FTLD 22 Alzheimer's 35 controls	64 67 64	-	Normal hearing	Frontotemporal lobar degeneration Alzheimer's disease	Bilateral anterior TL atrophy (14) hippocampal atrophy (16), generalized cerebral atrophy (4)	MRI scans (11 FTLD, 18 Alzheimer's)	FTLD: impaired familiarity, REC of familiar voices, intact unfamiliar voice DISCR Alzheimer's: impaired familiarity, REC of familiar voices, impaired unfamiliar voice DISCR	FTLD + Alzheimer's: impaired familiar face familiarity, REC, and apperceptive face processing	N/A	*VBM analysis* Right anterior TL = voice, name and face REC, Right inferior PL (angular gyrus) = unfamiliar voice DISCR
Roswandowitz et al., 2018	58 patients focal brain lesions 31 RBD 27LBD	48	46	Normal hearing (covariate in VLSM analysis)	ischemic stroke (34), intracerebral haemorrhage (6), subarachnoid haemorrhage (6), TBI (7), tumor (4)	Bilateral TL and right inferior PL well covered by lesions	MRI scans (56), CT scans (2)	Worse performance in voice-name test in patients compared to controls (Roswandowitz et al., 2014) RBD worse in voice REC of recently familiarised voices than LBD 9% of patients report poor voice REC after lesion onset	5% of patients report poor face REC after lesion onset	No severe aphasia	*VLSM analysis* Right mid/posterior TL = selective voice REC deficit Right inferior PL = impaired voice-face integration

All brain-behaviour findings rely on descriptive brain-behaviour associations if not stated otherwise.
MSO= Months since onset, LBD = Left-brain damaged patients, RBD = Right-brain damaged patients, BBD = Bilateral-brain damaged patients, N/A = not available, TBI = traumatic brain injury, LH = left hemisphere, RH = right hemisphere, DISCR = discrimination, REC = recognition, TL = temporal lobe, PL = parietal lobe, STS/G = superior temporal gyrus/sulcus, FFA = fusiform face area, FTLD = frontotemporal lobar degeneration, VBM = Voxel-based morphometry, VLSM = Voxel-based lesion symptom mapping

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Apperceptive voice-identity processing in acquired phonagnosia

Although the term phonagnosia was first mentioned in 1982 by van Lancker and Canter, examinations on voice-identity processes had begun almost a decade previously. These first studies addressed mainly *perceptual* aspects of voice-identity processing (e.g. unfamiliar voice discrimination). In 1976 the Swiss neurosurgeon Par G. Assal and his colleagues (Assal *et al.*, 1976) published the first study on acquired phonagnosia. They investigated 47 patients with unilateral brain lesions, including 25 patients with lesions in the right hemisphere (right brain damaged or RBD) and 22 patients with lesions in the left hemisphere (left brain damaged or LBD) as well as 29 healthy age- and handedness-matched controls. This study was centred on three main questions: (i) Does a deficit in voice discrimination after brain damage exist? (ii) Is the voice-discrimination deficit associated with right hemispheric lesions? (iii) Are voice-identity and language processes dissociable mechanisms? The authors showed that patients with brain lesions performed significantly worse than healthy controls on discrimination tasks with unfamiliar voices. Participants were tested on discrimination between unfamiliar adult male, female, and children's voices as well as on discrimination among only unfamiliar female voices either speaking different languages or the same language (i.e. French) (Table 1). RBD patients performed significantly below controls on all three tests (i.e. based on Tuckey-Hayes statistics), whereas LBD patients only performed worse than controls when discriminating female voices speaking different languages. A direct statistical patient group comparison however was not conducted. This was the first indication that impaired apperceptive voice-identity processing exists after brain damage and that it might be predominantly a function of the right hemisphere. Although no information about precise lesion locations was available the authors attempted to localise right hemispheric lesions relevant for voice discrimination with a dichotic listening test. RBD patients performed worse in voice discrimination if voices were presented to the left in comparison to the right ear. The authors speculated that voice discrimination may be assigned to the right temporal lobe.

Further, addressing the relation between voice-identity and language processes, the authors directly compared voice-discrimination abilities between LBD patients with and without aphasia (i.e. speech and language disorder caused by brain damage predominantly to the language-dominant left hemisphere). Performance in the voice-discrimination tests was similar for aphasic and non-aphasic LBD patients. This was a first indication of the separability of voice-

 doi:10.20944/preprints201806.0280.v2

identity processing from language abilities. In RBD patients language abilities were not considered, probably because it is unlikely that aphasia occurs in RBD patients. But in RBD patients visual abilities were tested. This was done with a visual figure/ground discrimination task (Poppelreuter test) and a visual-spatial memory task. Results showed that unfamiliar voice-discrimination performance was significantly worse in RBD patients with impaired visual processing than in RBD patients with intact visual processing. Whether the RBD patients with intact visual processing had nevertheless voice-discrimination difficulties in contrast to healthy controls was not tested.

Five years later, Assal and colleagues (1981) elaborated on their pioneering study by assessing apperceptive voice mechanisms by testing voice-discrimination abilities alongside the ability to *detect* a change in speaker identity (Table 1). This time, Assal et al. investigated unfamiliar voice discrimination in a sample of 76 patients (40 LBD, 36 RBD) and 35 healthy controls and unfamiliar speaker-change detection in 52 patients (28 LBD, 24 RBD) and 11 healthy controls. The authors replicated their previous findings: (i) They found a right hemispheric dominance for apperceptive voice-identity processing. This time, the authors showed that RBD patients were impaired on *both* apperceptive voice tasks, i.e. unfamiliar voice discrimination *and* unfamiliar speaker change detection. Importantly, in contrast to the previous study, this time a direct statistical group comparison between RBD and LBD patients on voice discrimination yielded a significant group difference (ANOVA at $\alpha = 0.05$): RBD patients performed worse than LBD patients. (ii) Based on the dichotic listening results, the authors again suggested an important role of the right temporal lobe (this time more specifically of the temporo-parietal region) during voice discrimination. (iii) Again they noted that dissociation between speech and voice-identity processing was evident in this cohort; voice-discrimination performance was not different between aphasic and non-aphasic LBD patients.

The first case report of acquired phonagnosia: The case RB

Assal et al. (1981) also reported the first case study of acquired phonagnosia; case RB. He was a 45-year-old male, managing director, had musical training, and normal hearing abilities. After brain injury resulting from vascular disease, RB reported difficulties in music and irony perception, voice recognition, as well as speech and sound perception. While RB recovered from the latter two difficulties one month after lesion onset, he continued to evidence a strong

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

deficit in recognising familiar voices and a moderate deficit in discriminating voices compared to controls which was based on numerical group difference inspection. Unfortunately, details on the test designs were not reported. Interestingly, face recognition was tested and intact. A brain scan (not specified by the authors, but likely a CT scan) originally revealed bilateral cortico-subcortical lesions in the anterior temporal lobe, initially more pronounced in the right hemisphere that after six weeks resolved into a lesion predominantly in the left temporal lobe.

This first case report on acquired phonagnosia implicated a role for the temporal lobe in voice-identity processing and showed that voice-identity processing can be impaired while leaving face-identity processing intact. Further, the case report gave a first indication that voice recognition (associative voice-identity processing) and voice discrimination (apperceptive voice-identity processing) might be dissociable mechanisms.

In our view, it is remarkable that in these first studies Assal and colleagues asked questions that have traced all future studies on voice-identity processing. However, to date, these studies are relatively unknown in the field, probably because they are reported in French only.

Apperceptive and associative voice-identity processing in acquired phonagnosia

Van Lancker and colleagues took research on phonagnosia a decisive step further. Van Lancker and Canter (1982) investigated *associative* voice-identity processing in 30 patients with focal brain lesions (21 LBD/ 9 RBD) with a familiar voice-recognition test (Table 1). All LBD patients had aphasia. One aim of the study was to assess whether familiar voice recognition is primarily assigned to the right hemisphere, as found in the prosopagnosia (i.e. face-identity processing deficit) literature (De Renzi, 1986; Damasio *et al.*, 1990; De Renzi *et al.*, 1991). Further, van Lancker and Canter were interested in the relation between voice- and face-identity processing. Therefore, patients were tested on their voice- and face-recognition abilities. In both tasks, patients were asked to match a celebrity voice/face to a written name (Table 1). A deficit in the voice- and face-recognition task was more prevalent in RBD than in LBD patients. 4/9 RBD patients were impaired on familiar voice recognition. Only in 1 RBD patient this deficit was selective to voice recognition; the remaining three RBD patients also had a deficit in face recognition. In contrast, only 1/21 LBD patients had impaired familiar voice recognition and another one had impaired familiar face recognition. The authors concluded that associative

 doi:10.20944/preprints201806.0280.v2

voice-identity processing can be assigned to the right hemisphere. Further, the authors suggested that voice- and face-recognition deficits tend to co-occur and that both may rely on neuronal mechanisms within the right hemisphere. Further, the authors conclude that voice recognition might be dissociable from left-hemisphere language functions as 20 of 21 aphasic LBD patients had intact voice recognition. Interestingly, there were two cases in which voice-recognition impairments seemed to be selectively impaired, i.e. with intact face recognition (1 RBD, 1 LBD). The behavioural profile of these two patients might be indicative of specific neural mechanisms for familiar voice recognition that can be dissociated from those supporting language and face-recognition abilities.

Next, Van Lancker and Kreiman (1987) directly compared the relation between apperceptive and associative voice-identity processing by testing unfamiliar voice discrimination and familiar voice recognition in the same patients. Although both abilities were located in the right hemisphere in previous studies (Assal *et al.*, 1976; Assal *et al.*, 1981; Van Lancker and Canter, 1982), the case RB had indicated a potential dissociation between both mechanisms (Assal *et al.*, 1981). Van Lancker and colleagues tested 32 patients (15 LBD, 11 RBD, 6 bilateral brain damaged (BBD)) and healthy age- and education-matched controls (n = 48) on both unfamiliar voice discrimination *and* familiar voice recognition. All LBD patients had aphasia. In contrast to previous findings, patients with lesions in the left or right hemisphere were similarly impaired (relative to the control group; 2-way repeated measure ANOVA at $\alpha = 0.01$) in the unfamiliar voice-discrimination task. In contrast, only RBD patients showed impaired familiar voice recognition, as compared to controls (Van Lancker and Kreiman, 1987). LBD patients' familiar voice-recognition performance was similar to controls. Looking at the cases individually, 14 of the 32 patients showed dissociable behavioural performances in the unfamiliar voice-discrimination and familiar voice-recognition task (1 RBD, 6 LBD, and 3 BBD; no lesion lateralisation on the remaining 4 patients was reported). They had impaired voice discrimination and intact voice recognition or vice versa. Of the 10 patients for whom they reported individual results, worse voice discrimination was associated with LBD and worse voice recognition with RBD. BBD patients had both worse voice discrimination and recognition. Moreover, there was no correlation between the discrimination and recognition performance in patients. Taken together, these results suggested that both apperceptive and associative voice- identity processing might be underpinned by dissociable cognitive and neuroanatomical mechanisms.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

To assess the selectivity of a given voice-identity processing deficit, 4 patients with a severe deficit in either voice discrimination or recognition were also tested on their face- recognition and -discrimination abilities as well as environmental sound processing. A voice- specific deficit pattern emerged; face and sound processing was intact in those patients suggesting a fairly selective phonagnosia.

To reveal which anatomical regions within the respective hemisphere sub-serve apperceptive and associative voice-identity processing, van Lancker and colleagues (1988) studied 6 brain-lesioned cases for which CT scans were available. Patients were tested on both unfamiliar voice discrimination (apperceptive voice-identity processing) and familiar voice recognition (associative voice-identity processing) (Table 1). Patients' performance was compared to 30 healthy age-matched control participants. 5 of the 6 patients showed a clear discrepancy between the ability to discriminate unfamiliar voices and to recognise familiar voices (i.e. more than 2 SDs away from the controls' mean difference in test scores). Van Lanker et al. noted that the 3 patients who were exclusively impaired on unfamiliar voice discrimination had a lesion overlap in the temporal lobe of either the left or the right hemisphere (i.e. including anterior, mid and posterior regions) and were aphasic.

In contrast, the 2 patients with *selectively* impaired familiar voice recognition had in common lesions, which were located exclusively in the right hemisphere, including the posterior part of the temporal and parietal lobe structures such as the superior portion of the angular gyrus and the posterior supramarginal gyrus. The one patient who did not show dissociation between voice discrimination and recognition, being impaired on both tasks, had a lesion in the right mid/ posterior temporal lobe and the right parietal lobe including the superior angular gyrus and the supramarginal gyrus. The authors (Van Lancker *et al.*, 1988) discuss a relevant role of the bilateral temporal lobe for unfamiliar voice discrimination (apperceptive voice-identity processing) and of the lateral parietal lobe in the right hemisphere for familiar voice recognition (associative voice-identity processing) (Figure 2). This study provided supporting evidence for distinct mechanisms underlying apperceptive and associative voice-identity processing.

In a follow up study, Van Lancker *et al.* (1989) aimed to quantitatively confirm the descriptive behavioural and neuroanatomical dissociation between apperceptive and associative voice-identity processing. To allow a quantitative brain-behaviour analysis, they tested a large

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

sample of 56 brain-damaged patients. 44 patients (23 LBD, 15 RBD, 6 BBD) were tested on both an unfamiliar voice-discrimination and familiar voice-recognition task (Table 1). 12 patients (2 LBD, 10 RBD) were tested only on familiar voice recognition. All LBD patients were aphasic. Results were compared between lesion groups and healthy age- and education-matched control participants. Behavioural results showed that both LBD and RBD patients performed worse on the unfamiliar voice-discrimination task compared to controls (2-way repeated measure ANOVA at $\alpha = 0.05$). On the familiar voice-recognition task, only RBD patients were impaired, relative to controls. In line with previous findings (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988), unfamiliar voice discrimination (apperceptive voice-identity processing) was assigned to lesions in both the *left or right* hemispheres and familiar voice recognition (associative voice-identity processing) only to lesions in the *right* hemisphere. Next, they investigated the neuroanatomic substrates underlying this behavioural pattern. Based on 43 available CT scans, lesions were classified according to the lobe with the largest extend of the lesion. According to their hypothesis, a lesion in the right parietal lobe was significantly associated with a deficit in associative voice-identity processing (familiar voice-recognition task). All 9 patients with a right parietal lobe lesion showed impaired familiar voice recognition; as did 7 of 43 patients having the lesion elsewhere. Unfortunately, the authors did not report the lesion location of those 7 patients. It would have been interesting to observe whether lesions in these additional 7 cases were located adjacent to the right parietal lobe or in other regions such as the temporal lobe as suggested by Assal *et al.* (1981) and by neuroimaging findings (Figure 1 B).

The analysis of apperceptive voice-identity processing was based on 25 CT scans. Confirming the authors' hypothesis, 13 patients with a lesion in either the left or right temporal lobe performed worse in discriminating unfamiliar voices compared to controls. There were also 4 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. Of the patients having their lesion outside the temporal lobe, 9 had high and 4 had low scores on the discrimination task. Of these 4 patients with impaired voice discrimination, lesions were adjacent to the temporal lobe. In summary, van Lancker and colleagues provided quantitative evidence that lesions in the right parietal lobe were associated with associative voice-identity processing and lesions in either the left or right temporal lobe with apperceptive voice-identity processing (Figure 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Group evidence for selective voice-identity processing impairments

Previous studies were not conclusive as to whether phonagnosia may reflect a modality specific disorder. For example, while the case of RB (Assal et al., 1981) and 6 patients in van Lancker et al. (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987) suggested dissociation between voice- and face-identity processing, the patient group reported by van Lancker and Canter (1982) showed that voice- and face-identity deficits can co-occur. The same diversity emerged when considering the relation between voice-identity processing and other auditory processing abilities, such as speech, sound, emotion, and music processing (e.g., case RB, Assal *et al.*, 1981; and the 4 cases in Van Lancker and Kreiman, 1987).

To systematically assess the relation between voice-identity and identity processing of other sensory modalities as well as other auditory processes, Neuner and Schweinberger (2000) developed a comprehensive behavioural test battery. They studied 36 brain-lesioned patients (16 LBD, 13 RBD, and 7 BBD) for which brain surgery reports, CT or MRI scans were available and 20 healthy controls (matched in age, gender, and education). The test battery assessed apperceptive (discrimination tasks) and associative (familiarity decision and semantic association tasks) abilities of persons' voices, faces, and names (Table 1). In addition, the test battery included control tests on word, picture, and sound recognition to investigate the specificity of a given person-recognition deficit. In 13/36 patients, familiar voice recognition assessed by a familiarity decision task was significantly worse compared to controls' performance (cut-off for impairment: patient scores below the control mean at $\alpha = 0.05$ and 0.01). However, only 4 of the 13 patients showed a selective form of phonagnosia, with impaired familiar voice recognition, but intact sound, face, and name recognition. Unfortunately, for these cases, semantic association scores were not reported. 1 of these 4 patients also showed an overlapping impairment in voice discrimination; the lesion was located in the right hemisphere (Table 1). 2 of the 4 patients with selective familiar voice recognition deficits had a lesion located in the right hemisphere and one in the left hemisphere. Neuner and Schweinberger (2000) made large strides in investigating the specificity of phonagnosia. Their systematic investigation attested that phonagnosia can be witnessed as a specific deficit independent of nonverbal sound, face, and name recognition.

A study by Lang and colleagues (2009) specifically examined the relation between voice-identity and speech processing. In this study, familiar voice recognition was assessed (Table 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

The study included 20 brain-damaged patients (11 LBD, 9 RBD) and 17 healthy age-matched controls. The two groups were matched for lesion location and extent. Left-brain damaged patients were tested for aphasia (Aachen Aphasia Test). The results yielded a familiar voice-recognition deficit in RBD relative to performance in LBD patients and controls (one-factorial ANOVA at $\alpha = 0.05$). In contrast, LBD patients and healthy controls performed equally well on familiar voice recognition. The authors concluded that in LBD patients aphasia (5 amnestic, 5 Wernicke, 1 Broca aphasia) was not associated with familiar voice-recognition deficits. However, whether there is a double dissociation between voice-identity and speech processing remains open as language abilities were not assessed in RBD patients. Lesions in the right hemisphere were mostly confined to the supply areas of the middle cerebral artery and similar lesions in the left hemisphere did not affect familiar voice recognition. Unfortunately, more exact lesion location was not reported.

Case report evidence for selective voice-identity processing impairments

Hailstone *et al.* (2010) comprehensively evaluated voice-identity processing and several control tasks in 2 patients with neurodegenerative diseases (frontotemporal dementia) and 24 healthy age-matched controls. The authors assessed apperceptive (unfamiliar speaker-change detection) and associative (familiarity decision and semantic association) voice-identity processing as well as face, name, music, and sound processing (Table 1). Patient QR, 61-years old, had bilateral fronto-temporal atrophy, accentuated in the right anterior temporal lobe but extending posteriorly within the temporal lobe. Patient KL, 72-years old, had bilateral, predominantly anterior temporal lobe atrophy, which was more marked on the right hemisphere and in the inferior temporal cortices including the fusiform gyrus. In both patients, processing of familiar voices (familiarity decision, semantic association) was severely impaired in contrast to controls (modified *t*-test for single case studies at $\alpha = 0.05$; Crawford and Howell, 1998). In addition, both patients as compared to controls were impaired in familiar face and name processing. However, QR's face and name abilities were superior to KL's. This indicates a more selective phonagnosia in QR and a rather multi-modal person-identity processing deficit in KL. The person-identity processing deficits observed in QR and KL seemed to be restricted to associative processes. Apperceptive processing of voices (including perceptual processing of vocal-identity, vocal-gender, and speaker-size information) and faces were preserved in both. Hence, the authors classify the patients' deficits as associative agnosias. Both patients also

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

showed intact vocal-emotion recognition abilities. Processing of musical instruments in an auditory and visual task design however was affected in QR and KL. The authors suggested that the bilateral anterior temporal lobe is involved in supporting multiple aspects of person knowledge including voices, faces, and names with a right hemispheric dominance for aspects of nonverbal person knowledge.

Statistical brain lesion-behaviour relation: multimodal person recognition deficit

In the past decade sophisticated statistical approaches have been developed for high-resolution structural MRI group studies to afford more robust and objective associations between brain structure and behavioural performance (VBM: Ashburner and Friston, 2000; VLSM: Bates *et al.*, 2003). The first study assessing a statistical voxel-wise association between brain structure and voice-identity processing was published in 2011 by Hailstone and colleagues (Hailstone *et al.*, 2011). 36 patients with neurodegenerative diseases (14 Frontotemporal lobar degeneration (FTLD), 22 Alzheimer's disease) and 35 healthy controls (matched in age, gender, handedness, and education) were tested on a comprehensive behavioural test battery. For all 16 FTLD and 20 Alzheimer's disease patients, a high-resolution structural MRI scan was available. FTLD patients had atrophy in the anterior temporal lobes of both hemispheres. Of the Alzheimer's diseases patients, 16 had hippocampal atrophy and 4 patients had generalised cerebral atrophy. Participants were tested on apperceptive (unfamiliar speaker-change detection) and associative (familiarity decision and semantic association) voice-identity processing (see Table 1). To assess the selectivity of a given voice-identity processing deficit, within and across modalities, the test battery included tests on other measures of vocal processing (including speaker-size and vocal-gender information) as well as tests on face and name processing. In the associative voice tasks, both disease groups performed significantly worse compared to controls (z-tests and 95% Wald- type confidence intervals at $\alpha = 0.05$ and 0.001). However, the deficits were more profound in the FTLD than the Alzheimer's patients. A more heterogeneous pattern emerged for the apperceptive tests. During speaker-change detection and vocal-gender perception, only Alzheimer's patients were impaired. However, apperceptive face processing was impaired in both disease groups.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

By applying voxel-based morphometry, the authors presented neuroanatomical evidence that the anterior temporal lobe (predominantly of the right hemisphere), as well as the right fusiform gyrus, plays an important role in associative person recognition across different modalities, including voices, faces, and names (Figure 2). 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). For apperceptive voice-identity processing (speaker-change detection), the right inferior parietal lobe (i.e. angular gyrus) was found to be relevant (Figure 2). In light of the previous findings (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989), association of the parietal lobe with apperceptive voice processing is unexpected. However, based on patients' atrophy descriptions, lesions in the Hailstone *et al.* (2011) study covered mostly the anterior temporal lobes and thus results on parietal lobes might have to be interpreted with caution.

Statistical brain lesion-behaviour relation: selective voice-identity recognition deficit

In a recent study Roswandowitz *et al.* (2018) aimed to identify which lesion locations may cause a selective deficit in person-identity processing, which is confined to the auditory domain i.e. to voice-identity recognition. The authors were in particular interested in examining the contribution of the right inferior parietal lobe and the temporal lobe to voice-identity recognition (Fig. 1 B, Stage II). Based on the acquired phonagnosia cases described above (see section 'Apperceptive and associative voice-identity processing in acquired phonagnosia'), the right inferior parietal lobe is crucial for voice-identity recognition. Conversely, neuroimaging studies on neurotypicals have consistently identified recruitment of the temporal lobe during voice-identity recognition tasks (see section 'Model of voice-identity processing'). To resolve this discrepancy of regions critical for voice-identity recognition, Roswandowitz et al. conducted a voxel-based lesion-behaviour mapping study in a cohort of 58 patients with unselected unilateral focal brain lesions (31 RBD, 27 LBD patients) and high-resolution structural brain images. The study included a comprehensive behavioural test battery including recognition tasks of recently-familiarised, i.e. newly-learned (voice-name, voice-face association learning) and familiar voices (famous voice recognition) as well as visual (face-identity recognition) and acoustic control tests (vocal-pitch and vocal-timbre discrimination). Voxel-based lesion-symptom mapping (VLSM) analyses revealed a strong association between lesions in the right

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

mid/posterior temporal and right inferior parietal lobe and the recognition of both recently-familiarised and familiar voices. However, a selective voice-recognition deficit, that was independent of face-identity processing *and* acoustical analyses of voice-identity features such as pitch and timbre, was associated only with lesions in the right mid/posterior *temporal lobe*. This finding implicated an obligatory function for the temporal lobe to voice-identity processing, making it the most likely key structure of the core-voice system. In contrast, lesions in the right inferior parietal lobe were associated with reduced voice-identity recognition when voices were associated with a face. This finding is similar to the earlier van Lancker studies where lesions in the right inferior parietal lobe were associated with reduced performances in tasks where patients had to match a famous voice to a display of faces (and their names) (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). Thus, the right inferior parietal lobe might have a facultative role during voice-identity processing only when additional face information is available. The study by Roswandowitz et al. is the first to provide group evidence for an association between spatially well-defined brain lesions and selective voice-identity processing impairments (Fig. 2).

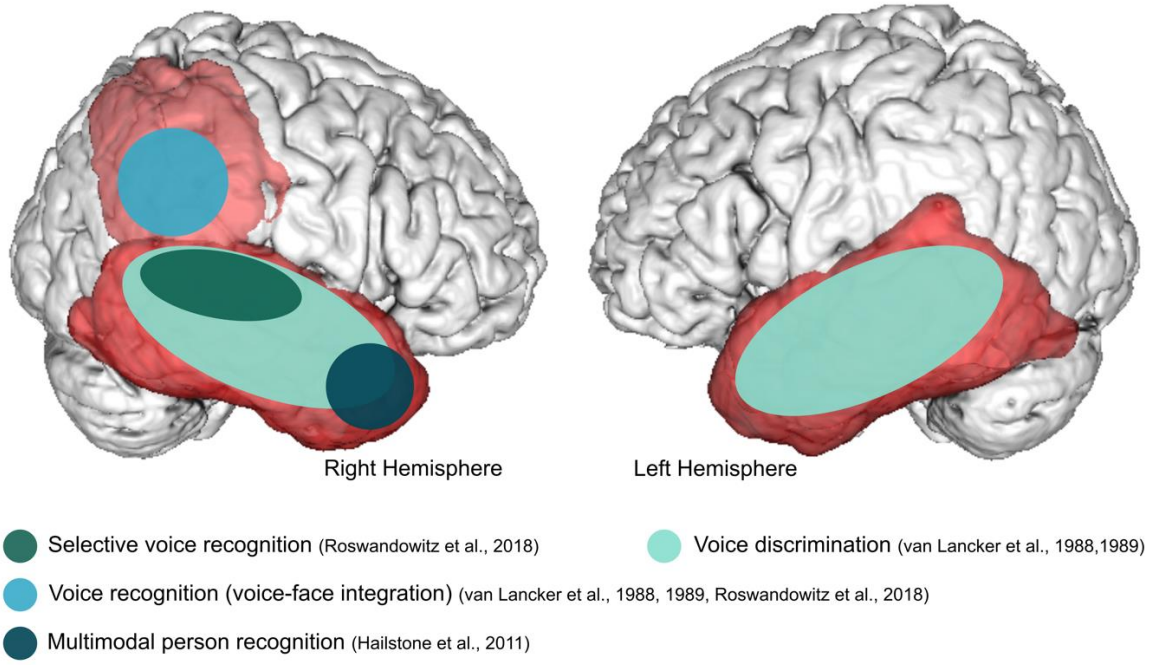

Figure 2. Schematic overview of studies reporting lesion locations associated with the respective voice-identity processing and multi-modal person recognition deficit. The temporal lobe is indicated by the dark red map and the right inferior parietal lobe by the light red map.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Developmental phonagnosia

Developmental phonagnosia has been discovered only recently (Garrido *et al.*, 2009; Herald *et al.*, 2014; Roswandowitz *et al.*, 2014; Xu *et al.*, 2015). Current prevalence estimates suggest that anything within the range of 0.2 % (Roswandowitz *et al.*, 2014), 1 % (Xu *et al.*, 2015) to 3.2 % (Shilowich and Biederman, 2016) of the population may have this deficit. While the precise aetiology of the deficit is unknown, it is possible that phonagnosia may have a heritable component, as has been observed in developmental prosopagnosia (Duchaine *et al.*, 2007; Grueter *et al.*, 2007; Schmalzl *et al.*, 2008; Lee *et al.*, 2010). In the following pages, we review the first documented cases of developmental phonagnosia, which have allowed for an examination of the nature and specificity of this developmental deficit.

The case of KH

Garrido and colleagues reported the first case of developmental phonagnosia, case KH (Garrido *et al.*, 2009). KH was a 60-year-old female who worked as a successful manager. She presented with a life-long impairment in voice recognition and reported that she failed to even recognise her daughter's voice on the phone. To confirm and assess the specificity of her self-report deficit, KH and a group of age-matched controls (n = 8) undertook a detailed behavioural battery of vocal-, visual-, and auditory-processing tests. As suspected, compared to controls, KH was significantly impaired in familiar voice-identity recognition. Specifically, her ability to judge whether a voice was famous or not was close to chance. This indicated weak feelings of familiarity towards known voice identities (i.e. impaired familiarity-association). In addition, her retrieval of names for the famous voices was negligible; KH could only accurately recall the name of one of the 48 presented famous identities, indicating impaired semantic-association. Her poor performance could not simply be explained by a lack of exposure to the vocal identities in every-day life[2]. When exposure to voices was explicitly controlled in a task, which required the learning of new unfamiliar speakers' voices with their corresponding name, KH's performance remained significantly poorer than age-matched controls (n = 8) for both naming and judging the familiarity (old/new judgment) of the speakers. Interestingly, KH's ability to discriminate between unfamiliar voices, that is to say whether two voice samples were articulated by the same

[2] In a post-test KH was asked to indicate if she had significant exposure in daily life to the voices, which she failed to name during testing. Taking this assessment into account, KH named only 3.85 % (i.e. 1 of 26) of the identities, which she stated she had significant exposure to.

 doi:10.20944/preprints201806.0280.v2

or a different speaker, was similar to controls under optimal listening conditions (task as described in Neuner and Schweinberger (2000), see Table 1). However, she was impaired in discriminating between identities when the task was made more difficult through the inclusion of auditory noise. When examining KH performance across tasks, the authors found no statistical evidence for dissociation between familiar voice recognition (associative voice-identity processing) and unfamiliar voice discrimination (apperceptive voice-identity processing).

Garrido and colleagues also examined whether KH's voice-identity processing impairment could be mediated by a higher-order multimodal person-recognition deficit, and, or a general deficit in vocal or auditory processing. Interestingly, KH's memory for faces was either superior to, or within the normal range of, controls. Her recognition and processing of general auditory information, including environmental sounds and musical excerpts, was normal; as was her ability to extrapolate vocal cues to support gender and emotion categorisation. In terms of speech processing, KH's performance was within the control range on a number of tasks, including vowel identification and the matching of verbal content to a visual target image. However, her performance under more challenging listening conditions was less clear. KH was impaired, relative to controls, in perceiving speech, which was embedded in auditory noise, although this impairment was not consistent across all levels of auditory noise. For example, KH's speech perception was impaired relative to controls for intermediate noise levels (SNR -3, SNR 3 dB), while her performance at the highest (SNR -6 dB) and lowest (SNR 6 dB) levels of auditory noise appeared normal. The authors attributed this poor performance to possible testing fatigue.

The case of KH suggested that developmental phonagnosia could represent a deficit in the processing of vocal identity, which was *not* mediated by a general deficit in the processing of auditory information, *nor* by a higher-level multimodal deficit affecting identity recognition across the visual and auditory domain. However, evidence for a possible dissociation between voice and speech processing, as well as voice recognition and voice discrimination, would become clearer in the following years, as more cases of developmental phonagnosia came to the attention of researchers (Roswandowitz *et al.*, 2014; Roswandowitz *et al.*, 2017).

 doi:10.20944/preprints201806.0280.v2

The case of AN

AN was a 20-year-old female university student who presented with a deficit in familiar voice recognition (Herald *et al.*, 2014; Xu *et al.*, 2015). Intriguingly, AN stated that she was not particularly aware of her deficit growing up as she had not thought that people could recognise an individual without seeing their face. Indeed, AN's face recognition was normal, as she obtained high scores on tests of familiar face recognition and naming (Xu *et al.*, 2015). Xu *et al.* (2015) and Herald *et al.* (2014) formally tested AN's familiar voice-recognition performance through a web-based experiment. In each trial, participants listened to samples of two voices; one celebrity and one non-celebrity voice. In parallel, 1, 2, or 4 celebrity face-name composites were presented. Participants first (i) decided, which of the two voices the celebrity voice was (i.e. familiarity decision) and then (ii) they indicated, which celebrity face-name composite matched the familiar rated voice. Relative to controls (n = 21, age range = 19-73 years), AN was markedly impaired in her ability to match the voices, which she classified as familiar with the correct celebrity face and name. Unfortunately, it was not explicitly reported whether her familiarity judgements towards the famous voices were also impaired. Conversely, AN's accuracy was similar to age-matched controls (n = 9) when the task was to choose, which of two unfamiliar voice samples matched a target voice. The target and test samples contained different verbal content. Given the dissociation between deficient familiar voice recognition and intact unfamiliar voice matching her behavioural profile is most likely indicative of an associative voice-identity processing impairment. Unfortunately, AN's abilities in other auditory tasks such as speech, emotion, and music processing were not formally assessed leaving open the possibility of additional impairments in other aspects of auditory processing.

The authors also examined the neuronal mechanisms underlying AN's voice-recognition deficit using two functional imaging experiments (Xu *et al.*, 2015). They employed (i) a standard functional localiser known to elicit voice-sensitive responses in the temporal voice areas (TVAs) of the STS/G (Belin *et al.*, 2000; see also Belin, Chapter 3, this edition) and (ii) a voice-imagery task. The study included AN and 9 controls (22-31 years). Functional imaging during the first experiment of passive listening to vocal, as compared to non-vocal, sounds (Belin *et al.*, 2000) demonstrated typical responses in AN in the TVAs, located bilaterally along the temporal lobes (Xu *et al.*, 2015). The second fMRI experiment assessed functional responses during voice imagery. Here, participants were presented with pictures of familiar persons' faces and names

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

and non-human object pictures and were asked to imagine the corresponding voice or sound after each image presentation. In a similar web-based test design, AN showed impaired imagery for famous voices, in comparison to, non-voice sounds[3]. Reduced blood oxygen level dependency (BOLD) responses were found in the ventromedial prefrontal cortex (vmPFC), left precuneus, and left cuneus in AN during voice, as compared to non-voice, imagery. The authors speculate that it is a dysfunction of the vmPFC, possibly driven by impaired fibre connections conveying voice information from the anterior temporal lobe to this region, that can explain AN's phonagnosia. However, unfortunately functional connectivity analyses have not been done in AN. A recent meta-analysis of neuroimaging studies on person recognition has revealed vmPFC involvement in famous person-identity processing independent of input modality (i.e., voice, face, and name), but not in identity processing of personally familiar or recently learned persons (Blank *et al.*, 2014). In our view, it is possible that atypical responses in the vmPFC may not fully explain AN's associative voice-identity processing for both personally familiar (based on a self-report) and famous voices. We speculate that the reduction of vmPFC responses in AN might be associated with her inability to imagine celebrity voices, but may not be causal for AN's phonagnosia.

AN's case suggested that typical responses in TVA of the STG/S for passive listening to voices (i.e. vocal, versus non-vocal sounds) may be observed in developmental phonagnosia. AN's intact matching of unfamiliar voices may have been supported by preserved TVA responses. Unfortunately, the integrity of the connectivity profile between the core-voice and extended system was not assessed in the case of AN, making it difficult to fully characterise the neural mechanisms of her associative voice-identity processing deficit. However, examination of two novel cases of developmental phonagnosia, reported by Roswandowitz *et al.* (2014), give rise to a more concrete understanding of how the core-voice and extended system may interact during voice-identity processing (Roswandowitz *et al.*, 2017).

[3] Xu and colleagues noted a similar pattern of low voice-imagery ratings for KH and also for SR, a 49-year-old male who also presented with poor voice recognition abilities (a full characterisation of the specificity of SR's phonagnosia was not reported). Interestingly, KH was not only impaired in voice imagery but also in non-voice imagery. However, the neurological underpinnings of KH and SR's behavioural deficits were not examined.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

The cases of AS and SP

Cases of developmental phonagnosia seem to be rare and have often come to the attention of researchers serendipitously (Herald *et al.*, 2014). Using a different approach, which involved large scale web-based testing of voice-recognition performance in ca. 1000 volunteers, Roswandowitz *et al.* (2014) identified two novel cases of developmental phonagnosia; AS and SP, both are successful academics. AS was a 32-year-old female, with no history of brain injury, who reported a distinct difficulty in voice-identity processing. For example, AS stated that she found it difficult to discriminate the voice of her daughter, from her daughter's friend, when they were playing in a nearby room. SP was a 32-year-old male who, like AS, reported a deficit in recognising speaker identity from the voice alone. Interestingly, SP only became explicitly aware that of his voice processing was atypical when his friend pointed out that a voice-over artist from their favourite show, which they watched together, had been replaced by a new vocal identity. This suggests that SP, unlike his friend, may have relied on compensatory strategies such as the use of current context, which remained unchanged in the case of the television show, to infer the identity of a voice.

Both AS and SP scored poorly on the original online web-based test designed to detect cases of phonagnosia (see Figure 3 for an illustration of the web- and laboratory-based screening measures used). This test assessed the ability to learn voice-name pairings for unfamiliar identities and subsequently recognise the learned voice by name. In a large comprehensive behavioural test battery, Roswandowitz *et al.* (2014) noted that this impaired association of vocal identities with additional semantic information was not limited to voice-name associations. Rather, relative to controls, both AS (controls n = 11) and SP (controls n = 10) scored poorly on tasks requiring the association of an unfamiliar voice with a colour or a facial identity. Interestingly, AS's performance on the unfamiliar voice-face learning task showed only a trend for impaired performance. Thus, AS may have had some preserved ability to use additional facial information to enhance the representation of the vocal percept. Both AS and SP showed normal face recognition performance as assessed by the Cambridge Face Memory Test (Duchaine and Nakayama, 2006), a standardised test used to detect cases of prosopagnosia and a novel unfamiliar face-name learning test (Roswandowitz *et al.*, 2014).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

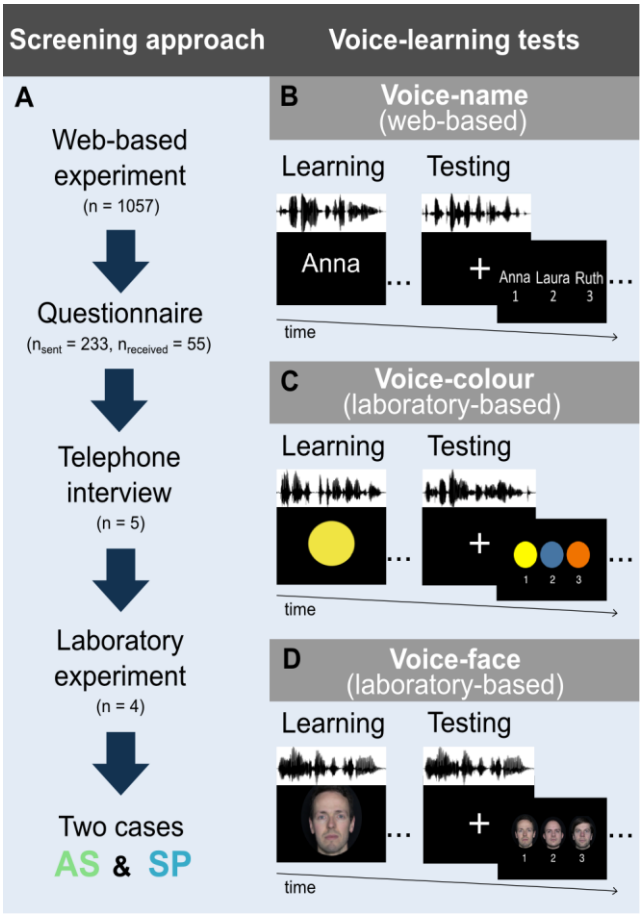

Figure 3. An overview of the web-based screening approach (A) and the voice-recognition tests (B-D) which were used to identify two unique cases of developmental phonagnosia, AS and SP (Roswandowitz *et al.*, 2014). Reprinted from Roswandowitz *et al.* (2014) with permission from Elsevier.

AS and SP's familiar voice recognition was also examined. Here, they were exposed to a series of famous and non-famous vocal identities. Following each voice sample, they were asked to indicate their familiarity with the voice and to provide a name or any uniquely identifying information pertaining to the vocal identity. Both phonagnosics showed atypical response strategies when classifying voices as familiar or unfamiliar, e.g. conservative (in AS) or liberal (in SP) rules. In contrast, only AS showed poor accuracy (*d'*) in categorising voices as familiar. Yet interestingly, she performed well in providing unique semantic information for the voices, which she successfully categorised as familiar. This pattern suggested that her voice-identity processing deficit was unlikely to be mediated by deficits at the level of semantic association. On the other hand, SP was poorer than controls in naming identities, which he classified as familiar to him, suggesting he may have an associative (semantic association) form of phonagnosia. This pattern was confirmed in an additional examination of unfamiliar voice discrimination were only

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

AS and not SP was significantly impaired relative to controls. As such, Roswandowitz and colleagues were the first to find evidence for a double dissociation between voice discrimination (apperceptive voice-identity processing) and recognition (associative voice-identity processing) in two cases of developmental phonagnosia. Both phonagnosics had normal hearing levels across a range of frequencies and both performed within the normal range on tests of speech processing in noise, music, and vocal-emotion recognition. Hence, their voice-identity processing impairments could not be attributed to a general deficit in auditory processing. However, both AS and SP were impaired on vocal-pitch perception. This impairment appeared to be voice specific, as neither phonagnosics were impaired on tests examining music pitch perception.

In a follow up study, (Roswandowitz *et al.*, 2017) examined the neural mechanisms underlying both AS and SP discrete voice-identity processing deficits. They firstly examined AS and SP's functional response profile in the core-voice system (see Figure 1A, B) using a vocal-sound experiment, where participants were exposed to a series of vocal and non-vocal sounds (Belin *et al.*, 2000). For SP, BOLD responses in the core-voice system were comparable to his controls (n = 16). This was in accordance with his associative phonagnosia where perceptual voice processing is intact. In contrast, AS's behavioural profile of poor perceptual voice processing (apperceptive phonagnosia) was mirrored in the reduced response in the core-voice system, specifically in the Heschl's gyrus compared to her controls (n = 14).

Secondly, the authors examined functional responses in AS and SP in a voice-identity recognition experiment. In this experiment, participants either performed a speaker or a speech recognition task on sentences spoken by different speakers (adapted from von Kriegstein *et al.*, 2003; Blank *et al.*, 2011; Schelinski *et al.*, 2016a). The authors observed that for the contrast speaker vs. speech task, AS showed reduced functional responses, relative to controls (n = 16), in regions of the core-voice system including the right antero-lateral Heschl's gyrus and planum temporale and extending to the right posterior STS/G. This finding is consistent with her apperceptive deficit. Conversely, AS had increased functional responses, relative to controls (n = 16), in the right temporal pole, and the right laterobasal amygdala - all proposed regions for the extended system (see Figure 1A, B). Interestingly, there was also a trend to significance for increased responses in the FFA in AS, which matches well with her relatively preserved ability to link voices with facial-identity information (voice-face learning test). Thus, it is likely that AS

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

uses additional facial information to enhance her weak perceptual processing of voices. Responses in the FFA are likely reflective of this cross-modal compensation (von Kriegstein *et al.*, 2006a; von Kriegstein *et al.*, 2008).

In contrast, in SP, connectivity *between* the core-voice and extended system was altered. As such, his deficit in voice-identity processing was likely to arise within the context of poor connectivity, rather than dysfunctional recruitment of the core-voice or extended system. In addition, SP showed increased response in, and increased functional connectivity within, the core-voice system during speaker (in contrast to speech) recognition. The authors propose that SP may rely more on the perceptual analysis of the voice to compensate for his associative phonagnosia. Enhanced recruitment of the core-voice system may be reflective of this.

Roswandowitz *et al.* (2017) findings were the first to show that responses in and connectivity between distinct brain regions can be associated with discrete behavioural subtypes of phonagnosia. Their findings demonstrated that cases of phonagnosia, which are associative in nature, may be marked by poor propagation of signals *from* the (intact) core-voice *to* the extended system (case of SP). Additionally, cases of apperceptive phonagnosia may be characterised by atypical functioning within the core-voice system itself (case of AS).

Identifying cases of developmental phonagnosia: Currently available methodology

Identifying cases of developmental phonagnosia can prove challenging. For example, the implementation of *standardised* screening tools for developmental phonagnosia, in comparison to prosopagnosia (see e.g. Duchaine and Nakayama, 2006; Duchaine *et al.*, 2007), is difficult. Unlike tests for face processing, tests for voice-identity processing are often constrained by the language of the listener, making testing beyond a geographical language location with the same vocal stimuli difficult.

Recently, attempts have been made to overcome such language constraints with the launch of the Glasgow Voice Memory Test (GVMT; Aglieri *et al.*, 2017), a brief test, which examines voice-identity processing. Specifically, in this test, listeners are exposed to a series of unfamiliar vocal identities uttering a single vowel and a series of unique bell sounds. Listeners

 doi:10.20944/preprints201806.0280.v2

must then immediately decide if these learned vocal identities, presented among a series of category matched distractor sounds, were present during the learning stage. The strength of this test is that the vocal stimuli are delivered as vowels, rendering language dependency minimal. Moreover, the bell condition permits for an assessment of the specificity of a voice-identity processing deficit. Aglieri *et al.* (2017) noted that the GVMT was sensitive in characterising KH as phonagnosic. Specifically, KH's scores for voice, rather than bell, recognition were significantly poorer than controls.

However, as mentioned throughout, findings from the acquired and developmental literature highlight voice-identity processing as a multistage process. Abnormalities arising during different stages of processing likely characterise the heterogeneity and subtypes of phonagnosia (Roswandowitz *et al.*, 2014; Roswandowitz *et al.*, 2017) (see Figure 1A). Thus, while the GVMT may offer a promising, globally available, screening tool for voice processing, it only assesses whether a general sense of familiarity is present for recently learned vocal identities. For example, we noted that SP, characterised by Roswandowitz *et al.* (2014) as a semantic-associative developmental phonagnosic, performed within the normal range on the GVMT for both voice and bell recognition. Indeed, Roswandowitz and colleagues had previously noted that SP's sensitivity towards voice familiarities was relatively unimpaired. Rather, he was poor at associating the familiar voice with identity specific semantic information. However, AS, an apperceptive developmental phonagnosic, was significantly impaired on both the voice and the bell recognition task (see Figure 4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

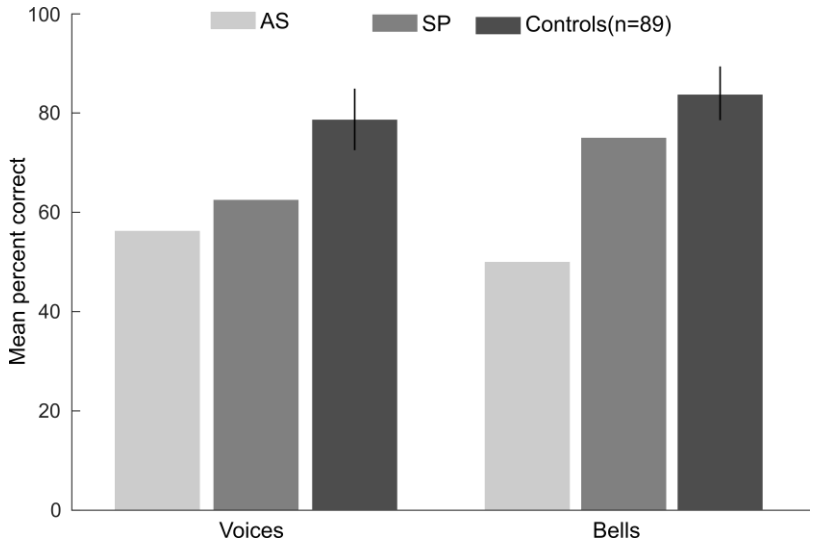

Figure 4. Plot showing AS and SP's voice and bell recognition performance on the GVMT, in relation to the 89 controls (31-40 years old) published by Aglieri et al. (2017). AS's voice and bell recognition were statistically different from controls; SP's performance on both tasks was comparable to controls (AS: voice: $p = 0.04$, bell: $p = 0.002$ SP: voice: $p = 0.14$, bell: $p = 0.40$). Statistical differences were assessed by comparing AS and SP's scores to control participants using a modified *t*-test (Crawford and Howell, 1998). *p* values are reported based on two-tailed probability, however the same pattern is evident for a one-tailed probability analysis. Error bars show 1 standard deviation of the mean.

Notwithstanding the challenge of differences in language, it is important that tests for phonagnosia are designed to address the multistage nature of voice processing; tackling voice perception, familiarity decision and semantic association (see Figure 1A and Table 1). It is possible that standardising the test design and testing procedure may allow for comparison across study findings for phonagnosia. Furthermore, it may also allow for a deeper insight into individual differences in voice-identity processing, at multiple stages, in the general population.

Phonagnosia in relation to current voice-identity processing models

Voice-identity processing as a multistage process: Supporting role of core-voice and extended system

Findings from neurotypical populations (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; Bestelmeyer *et al.*, 2011; Pernet *et al.*, 2015) and the acquired and developmental phonagnosia cases reviewed throughout support the temporal lobe as a key structure in the core-voice system implicated in the stage of perceptual voice-identity analysis and voice-identity recognition. For

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

instance, lesions in the mid/posterior temporal lobe are associated with a selective impairment in voice-identity recognition (Roswandowitz *et al.*, 2018).

What brain regions support specific stages of the voice-identity processing model? Lesions in the temporal lobe of the core-voice system, predominantly in the right hemisphere, have been consistently associated with impaired apperceptive voice-identity processing abilities, (Assal *et al.*, 1976; Assal *et al.*, 1981; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). Unfortunately, detailed lesion descriptions were not reported in these cases. In the case of AS, a developmental apperceptive phonagnosic, atypical responses in the auditory cortex (Heschl's gyrus) and posterior part of the temporal lobe (STG) were found (Roswandowitz *et al.*, 2017). Together, these findings highlight the importance of the temporal lobe, in particular the posterior part, and auditory cortex, in supporting the perceptual analysis of the voice (Figure 1A, Stage I, Figure 1B).

Based on findings on neurotypical populations the next stage of voice-identity processing, i.e. voice-identity recognition (Figure 1A, Stage II) is likely supported by the anterior/mid part of the right temporal lobe (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Andics *et al.*, 2010) within the core-voice system. The two lesion studies, which have assessed familiarity decision confirm the right-hemispheric involvement in familiarity decisions, but unfortunately do not provide detailed lesion descriptions (Neuner and Schweinberger, 2000; Lang *et al.*, 2009).

The subsequent stage of associative voice-identity processing, i.e. semantic association, may be supported by interactions between the core-voice and the extended system (Figure 1A, Stage III). Studies on acquired and developmental phonagnosia support the view that the anterior temporal lobe (Hailstone *et al.*, 2010, 2011; Roswandowitz *et al.*, 2017), the amygdala (Roswandowitz *et al.*, 2017) and the vmPFC (Xu *et al.*, 2015) may serve as potential candidates of the extended system (for review see Blank *et al.*, 2014). Currently, the evidence that semantic-associative phonagnosia may result from dysfunctional connections between the core-voice and extended system rests on the fMRI findings from the case of SP (Roswandowitz *et al.*, 2017). No other study has to date assessed the structural integrity or functional connectivity between these systems in phonagnosia cases. Note that direct damage to the extended system, rather than altered connectivity *between* the core-voice and extended system, would likely result in a multi- modal (i.e. non-voice selective) person-recognition disorder.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

Dissociations between stages of the voice-identity processing model

Findings from patients suffering brain damage (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988), neurodegenerative diseases (Hailstone *et al.*, 2010; Hailstone *et al.*, 2011), individuals with ASD (Schelinski *et al.*, 2016b), and cases of developmental phonagnosia (Roswandowitz *et al.*, 2014; Roswandowitz *et al.*, 2017) suggest a double dissociation between apperceptive and associative voice-identity processing abilities. There are reports on intact perceptual voice-identity analysis (Figure 1A, Stage I, apperceptive processing) and impaired familiarity decision and semantic association (Figure 1A, Stage II, III, associative processing) and vice versa. Also different lesion locations have been associated with impaired apperceptive and associative processes respectively.

In contrast, there is to-date only limited evidence for a double dissociation between the two stages of associative voice-identity processing — familiarity decision (Figure 39.1A, Stage II) and semantic association (Figure 1A, Stage III). Only a few studies compared both abilities intra-individually in cases with presumed associative voice-identity processing deficits (Garrido et al., 2009, Hailstone et al., 2010, 2011, Roswandowitz et al., 2014). Most of these cases showed marked overlapping impairments in familiarity decision and semantic association (Garrido et al., 2009, Hailstone et al., 2010, 2011, but see case SP, Roswandowitz et al., 2014).

Phonagnosia: Modality specific and cross-modal interactions

Acquired and developmental cases of phonagnosia confirm dissociable processing streams for person identification by voices, face, and names (Assal *et al.*, 1981; Van Lancker and Kreiman, 1987; Neuner and Schweinberger, 2000; Roswandowitz *et al.*, 2018). However, *interacting* mechanisms primarily between voice- and face-identity have been found as well. For instance, in acquired phonagnosia a deficit in voice- and face-identity processing tends to co-occur (Van Lancker and Canter, 1982; Hailstone *et al.*, 2010; Hailstone *et al.*, 2011). This could either suggest that the lesions commonly affect regions that process voice and face information independently, or alternatively it could suggest that there is a possible overlap in the neuroanatomical mechanisms, which support face and voice processing. Interestingly, cross-modal interactions between the face and voice regions have been observed in neurotypical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2018 doi:10.20944/preprints201806.0280.v2

populations (e.g. von Kriegstein *et al.*, 2008; Blank *et al.*, 2011). These interactions appear to be behaviourally relevant as voice recognition is often enhanced when the speaker has been previously learned by face (von Kriegstein and Giraud, 2006; Schweinberger *et al.*, 2007; O'Mahony and Newell, 2012; Schall *et al.*, 2013). Taken together, these findings suggest that phonagnosia can be modality specific. However, for potentially facilitating purposes, some degree of overlap, and or interactions between, the processing of faces and voices in both the typical and atypical brain is evident (for review see Maguinness and von Kriegstein, 2017).

Phonagnosia: Relations within the auditory modality

Phonagnosia findings also confirm separate pathways within the auditory modality. Cases of acquired and developmental phonagnosia have been described with intact speech and vocal-emotion recognition (Assal *et al.*, 1976; Assal *et al.*, 1981; Garrido *et al.*, 2009; Lang *et al.*, 2009; Hailstone *et al.*, 2010; Roswandowitz *et al.*, 2014). However, although such cases have been identified the findings are far from homogeneous. For example, in lesion studies, LBD patients with aphasia showed both intact (Assal *et al.*, 1976; Assal *et al.*, 1981; Lang *et al.*, 2009) and impaired (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989) voice-identity processing. Also in developmental phonagnosia, cases with impaired (Garrido *et al.*, 2009) and intact (Roswandowitz *et al.*, 2014) speech processing have been described. In neurotypical populations, interacting mechanisms between voice-identity and speech processing have been proposed. For example, voice-identity processing is facilitated when the speaker's language is familiar, rather than unfamiliar, to the listener (Perrachione *et al.*, 2011; Bregman *et al.*, 2012; Fleming *et al.*, 2014). Also speech content is more easily recognised when the speaker is familiar, rather than unfamiliar, to the listener (Nygaard and Pisoni, 1998; Levi *et al.*, 2011). Currently, whether in phonagnosia voice-identity and speech mechanisms do interact or are dissociable remains somewhat elusive. Anecdotal reports of phonagnosic participants suggest that they can rely on the way of speaking and speech content to recognise voice identity. We speculate that in phonagnosia, for instance, information about the speech content (e.g. reference to a past common situation) could help to recognise a speaker's voice identity. On the other hand, the facilitative effect of voice familiarity for speech recognition might not be available and could potentially explain difficulties with speech recognition in phonagnosia in certain cases, e.g. in KH for speech in noise (Garrido *et al.*, 2009). Such interactions would be in accordance with recent suggestions on potential neural mechanisms for

 doi:10.20944/preprints201806.0280.v2

interaction between speech and voice processing (von Kriegstein *et al.*, 2010; Kreitewolf *et al.*, 2014).

Conclusion

In sum, the present review suggests that voice-identity processing may represent a unique cognitive process, or more aptly process*es*, which can be selectively impaired. These processes appear to be supported by an interactive brain network. Disturbances arising at different stages of processing along this network, either due to brain insult or atypical development, may give rise to distinct impairments in the *apperception* and *association* of vocal identities. This heterogeneous behavioural profile emphasises the need for standardised behavioural testing, which takes into consideration the multistage nature of voice-identity processing and the resulting subtypes of phonagnosia. Furthermore, this multistage framework highlights that imaging studies should strive to address the integrity of connectivity *between* regions of the voice processing network, as well as responses within the regions themselves. We propose that the present classification of test designs (Table 1) could become a useful guide for future studies investigating voice-identity processing. Recent findings from the field of face processing suggest that both typical (Wilmer *et al.*, 2010; Zhu *et al.*, 2010) and atypical (developmental prosopagnosia) (Duchaine *et al.*, 2007; Grueter *et al.*, 2007; Lee *et al.*, 2010) face-identity processing may share a heritable component. These findings have been propelled by the availability of standardised testing designs for face perception and recognition. Using this same approach, it is possible that the coming years will provide insight into individual differences in voice-identity processing and their aetiology.

References

Aglieri V, Watson R, Pernet C, Latinus M, Garrido L, Belin P. The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. Behavior Research Methods 2017.

Andics A, McQueen JM, Petersson KM. Mean-based neural coding of voices. Neuroimage 2013; 79: 351-60.

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.

Assal G, Aubert C, Buttet J. Asymetrie cerebrale et reconnaissance de la voix. Review Neurology (Paris) 1981; 137(4): 255-68.

Assal G, Zander E, Kremin H, Buttet J. Discrimination des voix lors des lesions du cortex cerebral. Archives Suisses de Neurologie, Neurochirurgie et de Psychiatrie 1976; 119(2): 307-15.

Avidan G, Behrmann M. Impairment of the face processing network in congenital prosopagnosia. Frontiers in bioscience 2014; 6: 236-57.

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 PEG, 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.

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.

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.

Bregman MR, Creel SC, Creel S, Bregman MR. Learning to recognize unfamiliar voices: the role of language familiarity and music experience. 2012; 2012.

Crawford JR, Howell DC. Comparing an individual's test score against norms derived from small samples. The Clinical Neuropsychologist 1998; 12(4): 482-6.

Damasio AR, Tranel D, Damasio H. Face agnosia and the neural substrates of memory. Annual Reviews in Neuroscience 1990; 13: 89-109.

De Renzi E. Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. Neuropsychologia 1986; 24(3): 385-9.

De Renzi E, Faglioni P, Grossi D, Nichelli P. Apperceptive and associative forms of prosopagnosia. Cortex; a journal devoted to the study of the nervous system and behavior 1991; 27(2): 213-21.

Duchaine B, Germine L, Nakayama K. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. Cognitive Neuropsychology 2007; 24(4): 419-30.
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.
Ellis HD, Jones DM, Mosdell N. Intra- and inter-modal repetition priming of familiar faces and voices. BrJPsychol 1997; 88 (Pt 1)(Journal Article): 143-56.
Fleming D, Giordano BL, Caldara R, Belin P. A language-familiarity effect for speaker discrimination without comprehension. Proceedings of the National Academy of Sciences of the United States of America 2014; 111(38): 13795-8.
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.
Fox CJ, Iaria G, Barton JJS. Disconnection in prosopagnosia and face processing. Cortex 2008; 44(8): 996-1009.
Freud S. Zur Auffassung der Aphasien - Eine kritische Studie. Wien: Fischer Taschenbuch; 1891.
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.
Grueter M, Grueter T, Bell V, Horst J, Laskowski W. Hereditary prosopagnosia: the first case series. Cortex 2007; 43: 734-49.
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.
Herald SB, Xu X, Biederman I, Amir O, Shilowich BE. Phonagnosia: A voice homologue to prosopagnosia. Visual Cognition 2014; 22(8): 1031-3.
Kreitewolf J, Gaudrain E, von Kriegstein K. A neural mechanism for recognizing speech spoken by different speakers. Neuroimage 2014; 91: 375-85.
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).
Latinus M, Crabbe F, Belin P. Learning-Induced Changes in the Cerebral Processing of Voice Identity. Cerebral cortex 2011.
Latinus M, McAleer P, Bestelmeyer PE, Belin P. Norm-based coding of voice identity in human auditory cortex. Current biology: CB 2013; 23(12): 1075-80.
Lavner Y, Rosenhouse J, Gath I. The prototype model in speaker identification by human listeners. IJST 2001; 4(1): 63-74.
Lee Y, Duchaine B, Wilson HR, Nakayama K. Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. Cortex 2010; 46(8): 949-64.
Levi SV, Winters SJ, Pisoni DB. Effects of cross-language voice training on speech perception: whose familiar voices are more intelligible? J Acoust Soc Am 2011; 130(6): 4053-62.
Lissauer H. Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. Archiv f Psychiatrie 1890; 21(2): 222-70.

López S, Riera P, Assaneo MF, Eguía M, Sigman M, Trevisana MA. Vocal caricatures reveal signatures of speaker identity. Scientific Reports 2013; 3: 3407.
Maguinness C, Roswandowitz C, von Kriegstein K. Understanding the mechanisms of familiar voice recognition in the human brain Neuropsychologia 2018 (Special Issue: Familiar Voices).
Maguinness C, von Kriegstein K. Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia. Visual Cognition 2017: 1-14.
Neuner F, Schweinberger SR. Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn 2000; 44(3): 342-66.
Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. PerceptPsychophys 1998; 60(3): 355-76.
O'Mahony C, Newell FN. Integration of faces and voices, but not faces and names, in person recognition. Br J Psychol 2012; 103(1): 73-82.
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.
Perrachione TK, Del Tufo SN, Gabrieli JD. Human voice recognition depends on language ability. Science 2011; 333(6042): 595.
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.
Roswandowitz C, Kappes C, Obrig H, von Kriegstein K. Obligatory and facultative brain regions for voice-identity recognition. Brain 2018.
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 2017; 155: 97-112.
Schall S, Kiebel SJ, Maess B, von Kriegstein K. Early auditory sensory processing of voices is facilitated by visual mechanisms. Neuroimage 2013; 77: 237-45.
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.
Schmalzl L, Palermo R, Coltheart M. Cognitive heterogeneity in genetically based prosopagnosia: A family study. Journal of Neuropsychology 2008; 2(1): 99-117.
Schweinberger SR, Robertson D, Kaufmann JM. Hearing facial identities. QJExpPsychol(Colchester) 2007; 60(10): 1446-56.
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.
Shilowich BE, Biederman I. An estimate of the prevalence of developmental phonagnosia. Brain and Language 2016; 159: 84-91.
Stollhoff R, Jost J, Elze T, Kennerknecht I. Deficits in long-term recognition memory reveal dissociated subtypes in congenital prosopagnosia. PLoS One 2011; 6(1): e15702.
Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. Neuropsychologia 1987; 25(5): 829-34.
Van Lancker DR, Canter GJ. Impairment of voice and face recognition in patients with hemispheric damage. Brain and Cognition 1982; 1(2): 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.

von Kriegstein K. A multisensory perspective on human auditory communication. In: M. M. Murray MTW, editor. The Neural Bases of Multisensory Processes. Boca Raton: Taylor & Francis; 2011. p. 683- 700.

von Kriegstein K, Dogan O, Gruter M, Giraud AL, Kell CA, Gruter T, *et al.* Simulation of talking faces in the human brain improves auditory speech recognition. ProcNatlAcadSciUSA 2008; 105(18): 6747-52.

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, Giraud AL. Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. Cerebral cortex 2006a; 16(9): 1314-22.

von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD. How the human brain recognizes speech in the context of changing speakers. The Journal of neuroscience: the official journal of the Society for Neuroscience 2010; 30(2): 629-38.

von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD. Processing the acoustic effect of size in speech sounds. Neuroimage 2006b; 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.

Warrington EK. The selective impairment of semantic memory. Quaterly Journal of Experimental Psychology 1975; 27: 635-57.

Warrington EK, Shallice T. Category Specific Semantic Impairments. Brain 1984; 107: 829-54.

Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, *et al.* Human face recognition ability is specific and highly heritable. Proceedings of the National Academy of Sciences 2010; 107(11): 5238-41.

Xu X, Biederman I, Shilowich BE, Herald SB, Amir O, Allen NE. Developmental phonagnosia: Neural correlates and a behavioral marker. Brain Lang 2015; 149: 106-17.

Zhu Q, Song Y, Hu S, Li X, Tian M, Zhen Z, *et al.* Heritability of the specific cognitive ability of face perception. Current Biology 2010; 20(2): 137-42.