Hearing but Not Recognizing? Neural Mechanisms of Phonagnosia

Authors: Zhou Aibao, Hu Yanbing, Zhou Yingxin, Li Yu, Li Wenyi, Zhang Haobo, Guo Yanlin, Guoqing Hu

Date: 2020-11-04T00:00:00+00:00

Abstract

Voice identity recognition is crucial for many aspects of social communication. While most individuals can identify speakers based on their voices, patients with phonagnosia appear to have lost this ability. Phonagnosia refers to impairments occurring at different stages of voice identity processing, with main manifestations including acquired phonagnosia, developmental phonagnosia, and their respective subtypes. The brain regions damaged in patients with acquired phonagnosia mainly include the temporal lobe, Heschl' s gyrus, and temporal pole, whereas developmental phonagnosia is primarily associated with atypical responses in the right posterior superior temporal sulcus and impaired functional connectivity between the temporal lobe and amygdala. Future research may focus on aspects such as screening methods for phonagnosia, scope delineation, and cultural differences.

Full Text

The Neural Mechanism of Phonagnosia

ZHOU Aibao, HU Yanbing, ZHOU Yingxin, LI Yu, LI Wenyi, ZHANG Haobo, GUO Yanlin, HU Guoqing
(School of Psychology, Northwest Normal University, Lanzhou 730070)

Abstract: Human voice recognition is critical for many aspects of social communication. Most individuals can identify a person from their voice, yet people with phonagnosia appear to have lost this ability. Phonagnosia refers to deficits occurring at different stages of voice identity processing, encompassing acquired phonagnosia, developmental phonagnosia, and their subtypes. In acquired phonagnosia, damaged brain regions primarily include the temporal lobe, Heschl' s gyrus, and temporal pole. Developmental phonagnosia is associated with atypical responses in the right posterior superior temporal sulcus and

dysfunctional functional connectivity between the temporal lobe and amygdala. Future research may focus on screening methods, scope definition, and cultural differences in phonagnosia.

Keywords: acquired phonagnosia, developmental phonagnosia, subclassifications of phonagnosia, neural mechanism

The ability to identify a speaker from their voice seems innate, yet some individuals have lost this capacity. Van Lancker and Canter (1982) termed this specific impairment in voice identity processing "phonagnosia." Phonagnosia refers to a specific deficit in processing voice identity while largely preserving the ability to process other vocal information (such as gender, age, and emotion) as well as music and facial information (Neuner & Schweinberger, 2000; Roswandowitz et al., 2014; Stevenage, 2018). Currently, phonagnosia can be classified from two perspectives: first, based on the presence or absence of brain damage, it can be divided into acquired phonagnosia (voice identity processing deficits following brain injury) and congenital/developmental phonagnosia (voice identity processing deficits present from birth without brain damage) (Hailstone et al., 2011; Roswandowitz et al., 2014; Xu et al., 2015). Second, based on different stages of voice identity processing, it can be categorized into apperceptive phonagnosia and associative phonagnosia, which are also considered subtypes of phonagnosia. Specifically, apperceptive phonagnosia involves deficits in the perceptual processing stage of acoustic features while preserving semantic and familiar voice processing, whereas associative phonagnosia involves loss of semantic and familiar voice processing abilities while retaining basic voice perception capabilities (Gainotti, 2018; Muhl & Bestelmeyer, 2019; Stevenage, 2018; Wu et al., 2020).

Furthermore, research has shown that when individuals hear a voice, they first conduct low-level analysis in the primary auditory cortex (PAC), followed by deeper structural analysis that processes speech, emotion, and identity information through three independent neural pathways (P. Belin et al., 2004; Scott, 2019). However, speech and emotion information in voices are not completely independent; in other words, processing deficits in vocal speech and emotion can similarly affect the processing of speaker identity information (Wu et al., 2020). Due to the scarcity of phonagnosia cases in the literature, the behavioral characteristics and underlying pathological mechanisms of phonagnosia remain controversial. Therefore, reviewing previous literature can help deepen scholars' understanding of phonagnosia and provide a research foundation for domestic scholars regarding its neural mechanisms. Additionally, elucidating the relationship between phonagnosia and voice processing can help domestic scholars better understand its pathological mechanisms. This paper reviews the neural mechanisms of both types of phonagnosia and their subtypes, and proposes future research directions.

2. Brain Mechanisms of Acquired Phonagnosia

Neuroscientific models of identity recognition typically include a core system and an extended system, with evidence from facial identity recognition research (Blank et al., 2015; Lin et al., 2013). Voice identity recognition similarly relies on these two systems. Specifically, the core voice system located in the auditory cortex includes Heschl' s gyrus (HG), planum temporale (PT), superior temporal gyrus (STG), middle temporal gyrus (MTG), and superior temporal sulcus (STS) (Roswandowitz et al., 2017). The extended system includes the precuneus/posterior cingulate, temporal pole, amygdala, inferior frontal gyrus (IFG), and other channel-related brain regions such as the fusiform face area (FFA) (Aglieri et al., 2018; Blank et al., 2015; Schall et al., 2013). Research indicates that brain regions in the extended system are functionally and structurally connected to core voice system regions (Von Kriegstein & Giraud, 2006; Von Kriegstein et al., 2005). Could damage to brain regions related to these two systems lead to acquired phonagnosia? Researchers have conducted extensive studies and explored the neural mechanisms by examining acquired phonagnosia and its subtypes.

In studies of speaker discrimination and recognition, D. Van Lancker and Canter (1982) investigated 30 patients with unilateral brain damage, including 21 with left brain damage (LBD) and 9 with right brain damage (RBD). They found that RBD patients (vs. LBD) showed more prevalent familiar voice recognition deficits, suggesting that associative voice processing is related to the right hemisphere rather than the left. Subsequent studies showed that both LBD and RBD patients performed worse than healthy controls on unfamiliar voice discrimination tasks, while bilateral brain damaged (BBD) patients exhibited processing deficits in both voice discrimination and recognition (D. Van Lancker & Kreiman, 1987). Further research demonstrated that familiar voice recognition deficits were significantly correlated with right parietal lobe damage, while voice discrimination deficits were associated with temporal lobe damage in both hemispheres (Lancker et al., 1989; Roswandowitz et al., 2018; Van Lancker et al., 1988). Therefore, different brain-damaged patients likely show dissociations at different stages of voice processing, specifically apperceptive phonagnosia with deficits in voice perceptual processing stages and associative phonagnosia with deficits in voice recognition processing stages.

2.1 Acquired Apperceptive Phonagnosia

Acquired apperceptive phonagnosia refers to functional deficits in the perceptual processing stage of voice identity following brain injury, with behavioral abnormalities primarily manifested in the loss of acoustic feature perception and unfamiliar voice discrimination. First, regarding deficits in low-level auditory signal processing, age-related functional decline and partial brain atrophy may lead to dysfunction in the cognitive processing of basic acoustic features of voices. One study used MRI to investigate 20 Alzheimer' s disease patients with an average age of 66. Clinical diagnosis revealed that 16 patients had dis-

proportionate hippocampal atrophy and 4 had widespread brain atrophy. In a subsequent voice discrimination task requiring detection of speaker changes across a series of spoken phrases, Alzheimer' s patients showed deficits in low-level auditory signal analysis compared to healthy controls, specifically difficulty perceiving basic acoustic features such as pitch and timbre. Brain imaging results revealed that damage to the right inferior parietal lobe was closely related to apperceptive phonagnosia (Hailstone et al., 2011). Another study using repetitive transcranial magnetic stimulation (rTMS) provided direct evidence for the relationship between temporal voice areas (TVAs) and voice detection ability. Specifically, when rTMS targeted the right TVA, individuals' ability to discriminate between human and non-human voices was impaired (Bestelmeyer et al., 2011), demonstrating that the right TVA is crucial for low-level auditory signal processing of voices.

Second, in studies of deficits in voice perception stage processing, Papagno et al. (2018) recruited 29 patients who had undergone unilateral temporal lobe glioma resection and required them to perform unfamiliar voice discrimination tasks. Results showed that patients with right glioma resection (vs. left) exhibited deficits in voice perception stage processing. Subsequent fMRI brain scans supported previous findings of extensive damage to the right TVA and revealed that lesion areas extended to the right frontal and right posterior temporal lobes, including HG, fusiform gyrus, medial subcortical, and insular cortices. Additionally, the study suggested that disconnection between right temporal and frontal lobe functional connectivity might be the primary cause of apperceptive phonagnosia, providing substantial evidence for the importance of connections between core and extended voice systems in voice perception processing. A study examining voice discrimination across speaker identities including self, familiar, and unfamiliar individuals further confirmed the importance of the right hemisphere in voice perception stage processing (Candini et al., 2018).

Thus, although most studies indicate that apperceptive phonagnosia after brain damage is closely related to the right hemisphere, a few studies have also found that partial brain regions in the left hemisphere participate in voice perception stage processing. It remains uncertain whether damage to the temporal, frontal, or parietal lobes in the right hemisphere, or disconnection between different brain regions, could still cause voice recognition processing deficits, namely acquired associative phonagnosia.

2.2 Acquired Associative Phonagnosia

Acquired associative phonagnosia refers to functional deficits in voice identity recognition or semantic processing stages in brain-injured individuals, with behavioral manifestations primarily in the inability to recognize familiar voices or perform semantic association processing. In studies of speaker identity recognition and semantic association deficits, Hailstone et al. (2010) first used MRI to scan two Alzheimer' s patients. One patient showed bilateral anterior temporal lobe atrophy, including inferior temporal cortices containing the fusiform gyrus,

with more severe atrophy in the right hemisphere. The other showed bilateral fronto-temporal atrophy with aggravated right anterior temporal lobe (ATL) atrophy. Subsequent cross-modal matching experiments (e.g., voice-face matching, voice-name matching, sound perception tasks) revealed that both patients (vs. healthy controls) exhibited familiar voice recognition deficits. Although this study did not identify specific processing deficits for voice recognition (as deficits were also found in familiar face and name recognition tasks), it demonstrated the importance of bilateral anterior temporal lobes in multimodal (face, voice, and name) recognition processing.

Subsequent research recruited patients with two types of neurodegeneration—frontotemporal dementia and Alzheimer' s disease—and found that the former performed worse on voice identity recognition tasks. The study indicated that reduced gray matter volume in the right ATL and anterior fusiform gyrus was the main cause of recognition deficits for voices, faces, and names (Hailstone et al., 2011). Cosseddu et al. (2018) conducted cross-modal recognition tasks for famous faces, voices, and names in a 56-year-old patient with frontotemporal degeneration trends. The patient showed deficits in face and voice recognition but preserved name recognition ability. Brain imaging revealed atrophy in the right ATL. This study further dissociated recognition deficits across different modalities but did not identify specific behavioral features or neural mechanisms unique to phonagnosia.

Young et al. (2020) argued that although phonagnosia involves severe deficits in voice processing, it does not affect cross-modal cognitive processing. To investigate the specificity of voice recognition deficits, one study tested 58 unilateral brain-damaged patients from multiple experimental perspectives: auditory (familiar voice recognition and voice discrimination tasks), visual (face recognition tasks), and cross-modal (voice-face and voice-name matching tasks). Results identified the right posterior or middle temporal lobe as critical structures for voice identity recognition, while the right inferior parietal lobe, particularly the supramarginal gyrus, was recruited only when additional face identity processing was required (voice-face matching tasks) (Roswandowitz et al., 2018). Luzzi et al. (2018) used Positron Emission Tomography (PET) and MRI to study a patient with right anterior temporal lobe stroke. The patient performed normally on tests of visuospatial memory, executive function, visual perception and spatial skills, praxis, speech, and conceptual thinking, as well as on tests of basic musical ability, familiar phonological recognition, unfamiliar face and voice discrimination, and familiar face recognition. However, the patient showed specific deficits in famous voice recognition. Brain imaging revealed two small ischemic lesions in the right subcortical region involving the lenticular and caudate nuclei and the right temporal pole. This case further demonstrated specific behavioral manifestations and corresponding brain regions for acquired associative phonagnosia, clarifying the neural mechanisms underlying famous voice recognition and adding evidence for right hemisphere involvement in familiar voice recognition processing. Although right hemisphere damage is important for explaining the pathological mechanisms of acquired associative phonagnosia, the role of the left

hemisphere cannot be ignored. Papagno et al. (2018) required 29 patients who underwent unilateral temporal lobe glioma resection to perform famous voice recognition tasks, specifically evaluating familiarity of famous voices, identifying semantics in voices, and recognizing the speakers' names. Patients with left temporal lobe resection had difficulty identifying the names of famous speakers from their voices, possibly because left temporal lobe damage caused deficits in speech cognitive processing (Candini et al., 2018; McGettigan et al., 2017).

In summary, initial research on acquired associative phonagnosia found that damaged brain regions could also cause cross-modal recognition deficits. Subsequent studies incorporated various control tests, including acoustic feature perception (fundamental frequency and formants) from an auditory perspective and face and name recognition tasks from a visual perspective. These tests aimed to exclude the influence of voice perception stage processing and cross-modal recognition stage processing while providing strong evidence for the specificity of acquired associative voice processing deficits and their corresponding brain regions.

3. Brain Mechanisms of Developmental Phonagnosia

Developmental phonagnosia, also known as congenital phonagnosia, refers to voice identity processing deficits without brain injury. Currently, there is no unified data on the prevalence of developmental phonagnosia. Some studies suggest this rare condition affects approximately 0.2% to 1% of the population (Roswandowitz et al., 2014; Xu et al., 2015). Another study estimates the prevalence at 3.2%, which is closer to the prevalence of developmental prosopagnosia (Maguinness & Von Kriegstein, 2017; Shilowich & Biederman, 2016). Although the pathology of developmental phonagnosia remains unclear, like developmental prosopagnosia, it may have a genetic basis (Djouab et al., 2020; Maguinness et al., 2018; Maguinness & Von Kriegstein, 2017; Tian et al., 2019). Garrido et al. (2009) reported the first case of developmental phonagnosia in a patient with lifelong voice identity processing deficits who, by self-report, could not even recognize her daughter' s voice on the phone. To confirm and assess the specificity of her voice processing deficits, the patient and a group of healthy controls underwent a series of behavioral tests including visual and auditory channels. Results showed that the patient exhibited significant deficits in familiar voice recognition compared to healthy controls, indicating developmental associative phonagnosia. However, this study did not find dissociation between associative and apperceptive voice processing. Research has shown that developmental phonagnosia also has two subtypes: developmental apperceptive phonagnosia and developmental associative phonagnosia (Roswandowitz et al., 2014). Currently, the neural basis of developmental phonagnosia and its subtypes mainly includes the following aspects.

3.1 Developmental Apperceptive Phonagnosia

Developmental apperceptive phonagnosia refers to voice identity perception stage processing deficits without brain trauma, possibly present from birth, with behavioral abnormalities primarily in early auditory signal processing and unfamiliar voice discrimination. First, regarding deficits in acoustic feature processing, Roswandowitz et al. (2014) conducted a large-scale online voice recognition ability test with approximately 1,000 participants. The test required participants to learn new voices and then judge their familiarity. Results showed that two participants scored more than two standard deviations below others. Subsequent clinical diagnosis revealed normal hearing and no pathological brain abnormalities, leading to a preliminary diagnosis of suspected developmental phonagnosia. A series of follow-up tests revealed that one patient performed poorly on voice classification and pitch discrimination tasks, indicating deficits in low-level auditory processing stages, but showed preserved semantic information processing in voices, confirming developmental apperceptive phonagnosia.

Second, regarding voice identity perception stage processing deficits in developmental phonagnosia, Roswandowitz et al. (2017) used fMRI to further discover that this patient' s behavioral deficits in voice perception stage processing were reflected in reduced blood oxygen level-dependent responses in core voice system regions, specifically including the right antero-lateral Heschl' s gyrus, planum temporale (PT), and extending to the right posterior superior temporal sulcus/gyrus. However, functional responses in extended system regions such as the right temporal pole and amygdala, as well as the fusiform face area (FFA), were significantly increased. This study linked developmental apperceptive phonagnosia with its neural mechanisms, identified the specificity of its processing deficits, and suggested that deficits in low-level sound stage processing may affect subsequent voice perception stage processing. Additionally, it indicated that apperceptive phonagnosia patients can utilize additional cross-modal information to compensate for deficits in voice perception stage processing. Previous research also supports these brain regions: specifically, lateral Heschl' s gyrus contributes to encoding voice identity information, PT is more sensitive to voices (vs. non-voices) and shows sensitive functional responses to sudden changes in constant timbre (P. Belin et al., 2000; Bonte et al., 2014; Von Kriegstein & Giraud, 2006). In summary, developmental apperceptive phonagnosia is related to dysfunction in brain regions involved in early auditory processing stages, specifically including HG specialized for voice pitch and PT for timbre processing.

3.2 Developmental Associative Phonagnosia

Developmental associative phonagnosia refers to cognitive processing deficits in recognizing familiar voices and semantic information, possibly due to genetic factors without brain injury history, with behavioral abnormalities in ineffective recognition of familiar people' s voices and related semantic information.

Research on deficits in speaker recognition and semantic association stages found that a 20-year-old female showed deficits in familiar voice recognition. In her first fMRI experiment, bilateral temporal lobe TVAs were activated when listening to human (vs. non-human) sounds. In her second fMRI experiment evaluating functional responses during sound imagery, participants were presented with familiar faces and names along with object photos and asked to imagine corresponding sounds after each presentation. Results showed this patient had deficits in imagining famous voices, with reduced blood oxygen level-dependent (BOLD) responses in the ventromedial prefrontal cortex, left precuneus, and left cuneus (Herald et al., 2014; Xu et al., 2015). The study suggested that dysfunction in the ventromedial prefrontal cortex caused deficits in familiar voice recognition, while activation of temporal voice areas in the superior temporal gyrus/sulcus when hearing voices indicated preserved voice perception processing, thus explaining her classification as developmental associative phonagnosia. Some studies have questioned this result, suggesting that functional responses in the ventromedial prefrontal cortex may not fully explain associative voice identity processing for familiar and famous voices. The reduced ventromedial prefrontal response in this study might be related to the patient' s inability to imagine famous voices but may not be causally related to phonagnosia (Blank et al., 2015; Maguinness et al., 2018). However, another study offered a different explanation, proposing that voice recognition is achieved through voice individuating cues (VICs) that activate corresponding temporal voice areas, and that this developmental associative phonagnosia patient had familiar voice recognition deficits due to inability to extract relevant familiar identity cues (Biederman et al., 2018).

Further research identified specific deficits in developmental associative voice recognition processing. Specifically, this patient (vs. healthy controls) performed worse on familiar voice recognition tasks but similarly on face recognition, emotion recognition, and musical ability tests (Roswandowitz et al., 2014). Subsequent fMRI exploration of the neural mechanisms underlying voice recognition deficits in this developmental phonagnosia patient revealed that weakened connectivity between the core voice system in the right middle temporal gyrus or inferior temporal gyrus and the extended system in the lateral part of the amygdala was a key factor causing developmental associative phonagnosia (Roswandowitz et al., 2017). This study contributed primarily in two aspects: first, it provided behavioral manifestations and neural mechanisms for developmental phonagnosia and its two subtypes; second, it demonstrated the role of functional connectivity between core and extended voice systems in voice recognition processing. Some research results also support this finding, specifically showing that right fronto-temporal functional connectivity responses positively correlate with individual voice recognition ability (Aglieri et al., 2018; Bodin & Belin, 2020).

In summary, the neural mechanisms underlying developmental voice recognition processing deficits mainly include reduced functional responses in the ventromedial prefrontal cortex, left precuneus, and left cuneus, while weakened functional

connectivity between core and extended voice systems also affects individual processing at the voice recognition stage.

4. Comparison Between Developmental and Acquired Phonagnosia Based on Brain Mechanisms

The discovery of phonagnosia confirms the existence of different pathways in the auditory channel. Specifically, cases and neural mechanisms of acquired and developmental phonagnosia have been described from the perspective of two phonagnosia subtypes: apperceptive and associative. However, due to limited research on phonagnosia, comparing these two pathologically different types of phonagnosia (acquired and developmental) better clarifies their pathological similarities and differences (as shown in Figure 1 [Figure 1: see original paper]).

First, regarding the core voice system, both acquired and developmental phonagnosia involve functional deficits in the right temporal lobe and Heschl' s gyrus (HG) during voice perception stage processing. However, developmental phonagnosia, without brain injury, shows more specific brain regions such as the superior temporal gyrus. Both types involve the middle temporal gyrus during voice recognition stage processing, but acquired phonagnosia corresponds to HG damage, while developmental phonagnosia shows dysfunction in the inferior temporal gyrus.

Second, regarding the extended system, acquired and developmental phonagnosia show considerable differences in extended system brain regions corresponding to voice perception stage deficits. Specifically, acquired phonagnosia involves varying degrees of damage in the right frontal lobe, right inferior parietal lobe, and fusiform gyrus. Developmental phonagnosia shows less dysfunction in extended system brain regions but instead exhibits stronger functional responses in the right temporal pole, amygdala, and fusiform face area (FFA). Damage to the extended system in acquired associative phonagnosia mainly manifests as damage to a specific brain region that subsequently spreads to other areas, such as ischemic lesions in the right subcortical region spreading to the temporal pole. However, current research has not found atypical responses in the extended system in developmental phonagnosia.

Third, regarding functional connectivity between core and extended voice systems, acquired and developmental phonagnosia show different mechanisms in voice perception stage processing. Specifically, acquired phonagnosia primarily manifests as damage to temporal-frontal functional connectivity, while developmental phonagnosia involves dysfunction in connectivity between the temporal lobe and lateral amygdala. For voice recognition stage processing, brain damage in acquired phonagnosia concentrates in specific brain regions without direct evidence for damage to connectivity between the two systems, whereas developmental phonagnosia cases have found more evidence, specifically dysfunction in connectivity between the anterior temporal lobe in the core voice system and the amygdala and ventromedial prefrontal cortex in the extended system.

In summary, beyond core and extended voice systems, both types of phonagnosia may involve cross-modal compensation mechanisms. Specifically, acquired phonagnosia patients may recruit the parietal lobe in additional face processing during voice-face matching, while developmental phonagnosia patients may enhance functional responses in both systems through information such as faces and names, thereby alleviating voice identity processing deficits. Future comparative studies of the two phonagnosia types could, on one hand, observe the specific damaged or dysfunctional brain regions corresponding to different voice identity processing stages in patients to further explore the specificity of neural mechanisms and behavioral manifestations, which would better define phonagnosia. On the other hand, although cases of acquired and developmental phonagnosia have confirmed separate processing for identity recognition through voice, face, and name, research has discovered cross-modal interactions between prosopagnosia and phonagnosia, suggesting the possible existence of dual-channel agnosia (Young et al., 2020).

Figure 1: Cognitive model of phonagnosia (Roswandowitz, 2017)

5. Phonagnosia and Voice Processing Disorders

As previously described, the essence of phonagnosia is deficits in processing identity information, specifically the inability to discriminate and recognize speakers. Although phonagnosia involves processing specificity, it cannot be ignored that deficits in processing speech and emotion information in voices indeed affect the processing of speaker identity information.

5.1 Phonagnosia and Speech Information Processing Deficits

In voice recognition research, individuals can simultaneously utilize speech cues and paralinguistic cues in voices to judge speaker identity. In other words, voice is also the primary medium reflecting individual speech ability, and speech ability may directly affect speaker identity processing. Research has shown a relationship between dyslexia and voice speech information processing deficits (Gabrieli, 2009). Perrachione et al. (2011) assessed the ability of dyslexic and control groups (healthy participants) to identify native and non-native speakers. Both groups performed at near-chance levels (50%) on non-native speaker identification tasks, but the dyslexic group performed worse than controls on native speaker identification tasks. This study indicates that voice identity processing deficits occur when abstract representations of speech in voices cannot connect with acoustic information (due to unfamiliar stimuli, i.e., sounds from non-native speakers) or when individuals lack the ability for abstract speech representation (due to impaired native language representation, i.e., dyslexia). Based on this study, individuals with dyslexia likely have impaired ability for abstract speech representation, preventing speech from connecting with voice. Could deficits in the cognitive process of converting auditory signals to abstract speech representation also cause missing voice speech information?

Previous studies have found relevant evidence in post-stroke aphasia patients. Results support the neural mechanism that speech comprehension deficits cause voice speech information processing deficits, specifically that post-stroke aphasia patients' speech comprehension deficits likely relate to damage in the inferior temporal gyrus, fusiform gyrus, left posterior temporal region, and connectivity among middle temporal gyrus, inferior temporal gyrus, and cingulate cortex. Additionally, the posterior, lateral, and inferior temporal lobes may be integration centers for voice and abstract speech information (Bonilha et al., 2017). Subsequent research on acute aphasia patients further found that damage to the temporo-parietal area may be the main cause of sentence structure processing deficits (Kristinsson et al., 2020).

Thus, voice speech processing deficits are mostly related to damage in left temporal lobe brain regions, which seems to provide evidence for semantic association deficits in associative phonagnosia. Specifically, previous patients with left brain damage in associative phonagnosia studies mostly performed worse on speech-related stimulus tasks, suggesting that phonagnosia patients' deficits in voice semantic cognitive stage processing may relate to deficits in speech information cognitive processes. Furthermore, future research should focus on comparative studies between voice speech information processing deficits (such as dyslexia and aphasia) and phonagnosia to better define different categories of auditory agnosia.

5.2 Phonagnosia and Emotional Information Processing Deficits

Beyond perceiving identity and speech information from speakers' voices, emotional information in voices is ecologically and socially significant for listeners. In the absence of visual stimuli, individuals can infer emotional states from speakers' vocal tones, yet some groups may have lost this ability.

Regarding psychiatric populations, individuals with autism spectrum disorder typically show dysfunction in three aspects: restricted and repetitive behaviors, speech communication deficits, and social interaction problems. Consequently, compared to healthy individuals, autistic patients have difficulty understanding complex social and emotional signals expressed by speakers. Research has found that autistic patients show varying degrees of damage in the amygdala, superior temporal cortex (STC), and cerebellum—brain regions that belong to the core network for emotional voice processing (Fruhholz et al., 2016). Studies on schizophrenia patients have also identified deficits in processing emotional information in voices. Similar brain regions to those in autism patients have been found in schizophrenia groups, specifically including STG, amygdala, and auditory cortex dysfunction (Leitman et al., 2011). Additionally, missing functional connectivity is one of the main causes of abnormal emotional information processing in schizophrenia patients, specifically that weakened functional connectivity between the inferolateral cortex (IFC) and STC in schizophrenia patients leads to reduced perception of emotional clarity in voices (Leitman et al., 2010). Compared to schizophrenia patients, individuals with mood disorders

(such as major depressive disorder and bipolar disorder) also show processing deficits in vocal emotion cognition, primarily manifested as strong negative bias toward emotional information in voices and abnormal emotion regulation functions (Kanske & Kotz, 2012; Savitz & Drevets, 2009). In brain imaging studies, mood disorder patients (vs. healthy controls) showed reduced responses in STC, auditory cortex, amygdala, and IFC when hearing emotional voices. Research suggests that abnormal functional connectivity between the subgenual anterior cingulate cortex and limbic system may be one of the main causes of emotional information processing deficits in mood disorder patients (Connolly et al., 2013; Mitchell et al., 2004).

Second, brain-injured patients also show functional deficits when processing emotional voices, including stroke patients, brain atrophy patients, and surgically brain-injured patients (Fruhholz & Staib, 2017). Research on stroke patients found that right-hemisphere stroke patients lost almost all ability to perceive prosody in voices due to ischemic lesions in the supramarginal gyrus and associated white matter tracts (Patel et al., 2018). Research on brain atrophy has focused on dementia patients, emphasizing degeneration in the temporal lobe and partial frontal lobe. The atrophy process causes structural lesions in associated brain regions such as STC, IFC, medial frontal cortex (MTC), and amygdala, with damage in these regions almost always related to abnormal emotional voice processing (Omar et al., 2011; Rohrer et al., 2012). A study of epilepsy patients who underwent unilateral anterior temporal lobe resection found contributions of the amygdala and hippocampus to recognizing different types of vocal emotions (Khalfa et al., 2008).

In summary, although few studies have examined voice identity and emotion, deficits in vocal emotion processing across different patient types reveal that damage to right hemisphere brain regions, amygdala, and temporal lobe areas are the main causes of emotional information processing deficits. Some of these brain regions are closely related to the core voice system and extended system corresponding to voice identity processing. Future research should focus on the specificity of voice emotion and identity information processing and their corresponding brain regions.

6. Future Research Directions

This paper reviewed the neural mechanisms of two types of phonagnosia and their subtypes, finding that phonagnosia is related to lesions in the temporal lobe, Heschl' s gyrus, planum temporale, amygdala, temporal pole, parietal lobe, and functional connectivity between core and extended voice systems. Damage or dysfunction in different brain regions affects patients' voice identity processing deficits. Currently, controversies remain regarding the specificity of processing deficits and pathological mechanisms in phonagnosia patients. Future research could focus on the following aspects:

6.1 Phonagnosia and Screening Methods

Screening phonagnosia cases has proven challenging. Compared to face processing tests, voice identity processing tests are often limited by the language backgrounds of listeners and speakers, making testing across different language backgrounds difficult. Previous studies have attempted to address this through tests such as the Glasgow Voice Memory Test (assessing voice encoding and recognition ability) and the Bangor Voice Matching Test (assessing voice perception ability) (Aglieri et al., 2017; Muhl et al., 2018). However, phonagnosia cases demonstrate that voice identity processing is a multi-stage process, with abnormalities at different stages leading to different types of phonagnosia and subtypes. Therefore, these tests provide insufficient comprehensive assessment of individual voice identity processing ability. Future research should, beyond addressing challenges from different language backgrounds, emphasize the essence of phonagnosia—deficits in multiple stages of voice identity processing, including voice perception ability, sound familiarity judgment ability, and semantic association ability. Additionally, testing multiple stages of voice identity processing in healthy individuals can provide deeper understanding of individual differences in voice identity processing.

6.2 Phonagnosia and Scope Definition

The essence of phonagnosia is deficits at different stages of voice identity processing. Speaker identities in voice identity recognition tasks mainly include strangers, famous people, and acquaintances. Phonagnosia patients selected this way only demonstrate processing deficits for these three speaker identities. However, previous research has overlooked investigation of individual deficits in self-voice processing. Research has shown that self-voice processing also includes perception and recognition stages (Zhou et al., 2020). This indicates that "self" can also serve as a type of speaker identity, and "self" is more valuable for research than other speaker identities for three reasons: first, examining the relationship between self-awareness and phonagnosia from the perspective of self-consciousness to determine whether their psychological mechanisms share similarities or differences; second, investigating whether brain regions with abnormal neural mechanisms related to "self-disturbance" share common lesions with phonagnosia neural mechanisms; third, expanding and refining the conceptual definition of phonagnosia to determine whether individuals show processing deficits in self-voice perception and recognition stages.

Furthermore, some studies have found symptoms similar to phonagnosia in psychiatric populations. These cognitive deficits are strikingly similar to phonagnosia, and notably, the pathological mechanisms of these cognitive impairments remain controversial. For example, autism groups show not only behavioral abnormalities in voice identity processing but also demonstrated dysfunction in voice speech and emotion information and cross-modal processing. Additionally, autism groups show large individual differences, likely because many autism subtypes remain undefined. Therefore, future research should combine

multiple cognitive impairments and utilize brain imaging technology and experimental paradigms designed to dissociate these cognitive processing deficits, which would better clarify the boundaries between phonagnosia and other psychiatric disorders.

6.3 Phonagnosia and Cultural Differences

Current phonagnosia research has mostly focused on Western populations, with abnormal psychological, neural, and pathological mechanisms all derived from Western evidence. However, cultural factors have been proven in numerous studies to modulate individuals' cognitive processing of different stimuli. For example, Eastern and Western individuals show different self-construals due to cultural background differences, leading to dissociations in cognitive processing of familiar and intimate people. Specifically, cultural differences in phonagnosia manifest in two aspects: first, phonagnosia screening mostly occurs due to deficits in processing familiar speaker identities. However, different cultural backgrounds lead to different self-construals—independence and interdependence—so individuals from Eastern and Western cultural backgrounds may show dissociations in processing familiar speaker identities. Second, voice is the carrier of speech, and speech is the symbol of culture. Within broad Eastern and Western cultural contexts, many subcultures exist. Therefore, studying voice identity processing deficits in different subcultural backgrounds contributes to deeper understanding of phonagnosia and can further advance research on its pathological mechanisms.

6.4 Phonagnosia and Auditory Conditions

Currently, phonagnosia research remains in its initial stages, primarily due to the scarcity and difficulty of screening phonagnosia cases. However, if voice recognition difficulty is increased—for example, by conducting voice identity discrimination or recognition tasks under difficult auditory conditions—phonagnosia phenomena may also appear in healthy populations. Brain imaging research seems to have provided corresponding evidence (Alain et al., 2018). Currently, difficult listening conditions are achieved through three approaches: first, increasing background noise in voices, such as in a "cocktail party" scenario where a healthy individual attempts to identify a speaker' s speech, emotion, and identity information from several meters away. This cognitive process involves not only voice processing but also intentionally focusing attention on the speaker, selectively enhancing processing of that speaker's voice information while suppressing all other irrelevant sounds. Second, using filtering methods—specifically, using acoustic software to filter out partial frequencies from sounds, causing partial disappearance of acoustic features such as fundamental frequency, formants, and loudness, which significantly reduces clarity of speech information and emotional prosody in voices, thereby increasing cognitive processing difficulty. Third, increasing complexity of speech structures, such as semantically ambiguous words or syntactically complex sentences.

Thus, increasing listening difficulty is equivalent to raising the baseline level of voice processing difficulty, causing individuals with normal voice processing under ordinary listening conditions to potentially exhibit phonagnosia phenomena under difficult listening conditions. In terms of brain mechanisms, because difficult listening conditions consume substantial cognitive resources, corresponding brain regions show more activity than under normal listening conditions. Therefore, increasing listening condition difficulty to induce phonagnosia phenomena in healthy individuals and conducting comparative studies with phonagnosia patients' behavioral manifestations and corresponding brain regions should become a future research direction.

References

林菲菲, 陈旭, 周春霞, 马建苓, 冉光明. (2013). 面孔失认症的神经机制. 心理科学进展, 21(10), 1755-1762.

伍可, 陈杰, 李雯婕, 陈洁佳, 刘雷, 刘翠红. (2020). 人声加工的神经机制. 心理科学进展, 28(05), 752-765.

周爱保, 胡砚冰, 鲁小勇, 申莎, 关香丽, 陈大亮, 崔嘉澈. (2020). 我听故我在? 自我声音识别机制的探索. 心理科学, 43(03),

Aglieri, V., Chaminade, T., Takerkart, S., & Belin, P. (2018). Functional connectivity within the voice perception network and its behavioural relevance. Neuroimage, 183, 356-365.

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. Behavior Research Methods, 49(1), 97-110.

Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta analysis. Human Brain Mapping, 39(7), 2695-2709.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature, 403(6767), 309-312.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. Trends in Cognitive Sciences, 8(3), 129-135.

Bestelmeyer, P. E. G., Belin, P., & Grosbras, M.-H. (2011). Right temporal TMS impairs voice detection. Current Biology, 21(20), R838-R839.

Biederman, I., Shilowich, B. E., Herald, S. B., Margalit, E., Maarek, R., Meschke, E. X., & Hacker, C. M. (2018). The cognitive neuroscience of person identification. Neuropsychologia, 116, 205-214.

Blank, H., Kiebel, S. J., & Von Kriegstein, K. (2015). How the human brain exchanges information across sensory modalities to recognize other people. Human Brain Mapping, 36(1), 324-339.

Bodin, C., & Belin, P. (2020). Exploring the cerebral substrate of voice perception in primate brains. Philosophical Transactions of the Royal Society B, 375(1789), 20180386.

Bonilha, L., Hillis, A. E., Hickok, G., Den Ouden, D., Rorden, C., & Fridriksson, J. (2017). Temporal lobe networks supporting the comprehension of spoken words. Brain, 140(9), 2370–2380.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. The Journal of Neuroscience, 34(13), 4548–4557.

Candini, M., Avanzi, S., Cantagallo, A., Zangoli, M., Benassi, M., Querzani, P., ···Frassinetti, F. (2018). The lost ability to distinguish between self and other voice following a brain lesion. NeuroImage: Clinical, 18, 903–911.

Connolly, C. G., Wu, J., Ho, T. C., Hoeft, F., Wolkowitz, O. M., Eisendrath, S. J., ···Paulus, M. P. (2013). Resting-state functional connectivity of subgenual anterior cingulate cortex in depressed adolescents. Biological Psychiatry, 74(12), 898–907.

Cosseddu, M., Gazzina, S., Borroni, B., Padovani, A., & Gainotti, G. (2018). Multimodal face and voice recognition disorders in a case with unilateral right anterior temporal lobe atrophy. Neuropsychology, 32(8), 920–930.

Djouab, S., Albonico, A., Yeung, S. C., Malaspina, M., Mogard, A., Wahlberg, R., ···Barton, J. J. S. (2020). Search for face identity or expression: Set size effects in developmental prosopagnosia. Journal of Cognitive Neuroscience, 1–17.

Fruhholz, S., & Staib, M. (2017). Neurocircuitry of impaired affective sound processing: A clinical disorders perspective. Neuroscience & Biobehavioral Reviews, 83, 516–524.

Fruhholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. Neuroscience & Biobehavioral Reviews, 68, 96–110.

Gabrieli, J. D. E. (2009). Dyslexia: A new synergy between education and cognitive neuroscience. Science, 325(5938), 280–283.

Gainotti, G. (2018). How can familiar voice recognition be intact if unfamiliar voice discrimination is impaired? An introduction to this special section on familiar voice recognition. Neuropsychologia, 116, 151–153.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., ···Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. Neuropsychologia, 47(1), 123–131.

Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. Neuropsychologia, 48(4), 1104–1114.

Hailstone, J. C., Ridgway, G. R., Bartlett, J. W., Goll, J. C., Buckley, A. H., Crutch, S. J., & Warren, J. D. (2011). Voice processing in dementia: A neuropsychological and neuroanatomical analysis. Brain, 134(9), 2535–2547.

Herald, S. B., Xu, X., Biederman, I., Amir, O., & Shilowich, B. E. (2014). Phonagnosia: A voice homologue to prosopagnosia. Visual Cognition, 22(8), 1031–1033.

Kanske, P., & Kotz, S. A. (2012). Auditory affective norms for German: Testing the influence of depression and anxiety on valence and arousal ratings. Plos One, 7(1), e30086.

Khalfa, S., Guye, M., Peretz, I., Chapon, F., Girard, N., Chauvel, P., & Liégeois-Chauvel, C. (2008). Evidence of lateralized anteromedial temporal structures involvement in musical emotion processing. Neuropsychologia, 46(10), 2485–2493.

Kristinsson, S., Thors, H., Yourganov, G., Magnusdottir, S., Hjaltason, H., Stark, B. C., ···Rorden, C. (2020). Brain damage associated with impaired sentence processing in acute aphasia. Journal of Cognitive Neuroscience, 32(2), 256–271.

Lancker, D. R. V., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. Journal of Clinical & Experimental Neuropsychology, 11(5), 665–674.

Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P. D., & Javitt, D. C. (2010). Getting the Cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. Schizophrenia Bulletin, 36(3), 545–556.

Leitman, D. I., Wolf, D. H., Laukka, P., Ragland, J. D., Valdez, J. N., Turetsky, B. I., ···Gur, R. C. (2011). Not pitch perfect: Sensory contributions to affective communication impairment in schizophrenia. Biological Psychiatry, 70(7), 611–618.

Luzzi, S., Coccia, M., Polonara, G., Reverberi, C., Ceravolo, G., Silvestrini, M., ···Gainotti, G. (2018). Selective associative phonagnosia after right anterior temporal stroke. Neuropsychologia, 116, 154–161.

Maguinness, C., Roswandowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. Neuropsychologia, 116, 179–193.

Maguinness, C., & Von Kriegstein, K. (2017). Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia. Visual Cognition, 25, 644–657.

McGettigan, C., Jasmin, K., Eisner, F., Agnew, Z. K., Josephs, O. J., Calder, A. J., ···Scott, S. K. (2017). You talkin'to me? Communicative talker gaze activates

left-lateralized superior temporal cortex during perception of degraded speech. Neuropsychologia, 100, 51–63.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2004). Neural response to emotional prosody in schizophrenia and in bipolar affective disorder. British Journal of Psychiatry, 184(3), 223–230.

Muhl, C., & Bestelmeyer, P. E. G. (2019). Assessing susceptibility to distraction along the vocal processing hierarchy. Quarterly Journal of Experimental Psychology, 72(7), 1657–1666.

Muhl, C., Sheil, O., Jarutytė, L., & Bestelmeyer, P. E. G. (2018). The Bangor Voice Matching Test: A standardized test for the assessment of voice perception ability. Behavior Research Methods, 50(6), 2184–2192.

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain and Cognition, 44(3), 342–366.

Omar, R., Henley, S. M. D., Bartlett, J. W., Hailstone, J. C., Gordon, E., Sauter, D., ···Warren, J. D. (2011). The structural neuroanatomy of music emotion recognition: Evidence from frontotemporal lobar degeneration. Neuroimage, 56(3),

Papagno, C., Mattavelli, G., Casarotti, A., Bello, L., & Gainotti, G. (2018). Defective recognition and naming of famous people from voice in patients with unilateral temporal lobe tumours. Neuropsychologia, 116, 194–204.

Patel, S., Oishi, K., Wright, A., Sutherland-Foggio, H., Saxena, S., Sheppard, S. M., & Hillis, A. E. (2018). Right hemisphere regions critical for expression of emotion through prosody. Frontiers in Neurology, 9(224).

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. Science, 333(6042), 595–595.

Rohrer, J. D., Sauter, D., Scott, S. K., Rossor, M. N., & Warren, J. D. (2012). Receptive prosody in nonfluent primary progressive aphasias. Cortex, 48(3), 308–316.

Roswandowitz, C. (2017). Voice-identity processing deficit: The cognitive and neural mechanisms of phonagnosia (Unpublished doctorial dissertation), Humboldt University of Berlin.

Roswandowitz, C., Kappes, C., Obrig, H., & von Kriegstein, K. (2018). Obligatory and facultative brain regions for voice-identity recognition. Brain, 141(1), 234–247.

Roswandowitz, C., Mathias, Samuel R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. Current Biology, 24(19), 2348–2353.

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

Savitz, J., & Drevets, W. C. (2009). Bipolar and major depressive disorder: Neuroimaging the developmental-degenerative divide. Neuroscience & Biobehavioral Reviews, 33(5), 699–771.

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

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. Science, 366(6461), 58–62.

Shilowich, B. E., & Biederman, I. (2016). An estimate of the prevalence of developmental phonagnosia. Brain and Language, 159,

Stevenage, S. V. (2018). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. Neuropsychologia, 116, 162–178.

Tian, X., Wang, R., Zhao, Y., Zhen, Z., Song, Y., & Liu, J. (2019). Multi-item discriminability pattern to faces in developmental prosopagnosia reveals distinct mechanisms of face processing. Cerebral Cortex, 30(5), 2986–2996.

Van Lancker, D., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. Brain and cognition, 1(2), 185–195.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. Neuropsychologia, 25(5), 829–

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. Cortex, 24(2), 195–209.

Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. PLoS Biology, 4(10).

Von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. (2005). Interaction of face and voice areas during speaker recognition. Journal of Cognitive Neuroscience, 17(3), 367–376.

Xu, X., Biederman, I., Shilowich, B. E., Herald, S. B., Amir, O., & Allen, N. E. (2015). Developmental phonagnosia: Neural correlates and a behavioral marker. Brain and Language, 149, 106–117.

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. Trends in Cognitive Sciences, 24(5), 398–410.

Note: Figure translations are in progress. See original paper for figures.