Short review A "voice patch" system in the primate brain for processing vocal information?

Pascal C Belin, Virginia Aglieri

Hearing Research 366 (2018) 65–74

Contents lists available at ScienceDirect

Hearing Research

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

Short review

A "voice patch" system in the primate brain for processing vocal information?

Pascal Belin [a, b, *], Clémentine Bodin [a], Virginia Aglieri [a]

[a] Institut de Neurosciences de la Timone UMR 7289, Centre National de la Recherche Scientifique and Aix-Marseille Université, Marseille, France
[b] Département de Psychologie, Université de Montréal, Montréal, Canada

ARTICLE INFO

Article history:
Received 5 January 2018
Received in revised form
14 April 2018
Accepted 25 April 2018
Available online 7 May 2018

Keywords:
Voice
Conspecific vocalization
Category-selective cortex
Norm-based coding
Speaker identity
fMRI
Comparative approach

ABSTRACT

We review behavioural and neural evidence for the processing of information contained in conspecific vocalizations (CVs) in three primate species: humans, macaques and marmosets. We focus on abilities that are present and ecologically relevant in all three species: the detection and sensitivity to CVs; and the processing of identity cues in CVs. Current evidence, although fragmentary, supports the notion of a "voice patch system" in the primate brain analogous to the face patch system of visual cortex: a series of discrete, interconnected cortical areas supporting increasingly abstract representations of the vocal input. A central question concerns the degree to which the voice patch system is conserved in evolution. We outline challenges that arise and suggesting potential avenues for comparing the organization of the voice patch system across primate brains.

1. Introduction

There is currently renewed interest in studying non-human primates for understanding the evolution of speech and language, triggered by recent crucial pieces of evidence: e.g., that non-human primates show plasticity in their vocal production (Takahashi et al., 2015) (an ability thought reserved to songbirds or cetaceans), or that they have the anatomical ability to produce human vowels (Boe et al., 2017; Fitch et al., 2016) despite long-held belief (Lieberman et al., 1969). As summarized by Charles Snowdon in a recent commentary: "*Non-human primates do not talk, but we should not expect them to. Each species has its own adaptations for communication. Nevertheless there is much about language evolution that we can learn from non-human primates, provided that we study a variety of species and consider the multiple components of speech and language*" (Snowdon, 2017).

Here we focus on one often-neglected component of speech and language: voice perception. Speech after all consists of information carried by voice and so understanding the evolution of our ability to extract and process voice information is an integral part of the puzzle of language evolution. Indeed, before they started speaking and perceiving speech some tens of thousands years ago, our ancestors had lived for millions of years in an auditory environment rich in conspecific vocalizations (CVs), which presumably gave ample time for evolving neural mechanisms optimized for extracting different types of relevant information in CVs. To better understand the evolution of voice perception the *comparative approach*, based on comparison of perceptual and neural mechanisms between different extant species, is the method of choice: if cross-species similarities are high this could constitute evidence for homologous mechanisms inherited from a common ancestor, suggesting gradual evolution of voice perception. In contrast strong dissimilarities between humans and non-human primates would be evidence for abrupt changes (Fitch, 2000; Ghazanfar and Rendall, 2008; Rilling, 2014a, 2014b).

This paper considers the relatively recent evolution of voice perception in primates by briefly reviewing known perceptual and neural mechanisms of human voice perception and summarizing current equivalent knowledge in two other primate species: macaques (*Macaca mulatta*) and common marmosets (*Callithrix jacchus*) (Box 1). We conclude that although current evidence is

* Corresponding author. Institut de Neurosciences de la Timone UMR 7289, Centre National de la Recherche Scientifique and Aix-Marseille Université, Marseille, France.
E-mail address: pascal.belin@univ-amu.fr (P. Belin).

https://doi.org/10.1016/j.heares.2018.04.010

compatible with broadly similar voice perception mechanisms (with cortical "voice patches" observed in the three species), it is still too fragmentary for the in-depth comparisons necessary to understand the template for the primate voice patch system, and species-specific adaptations of that template. We end by listing some challenges that need to be overcome in future research into primate voice perception mechanisms.

Box 1
Why macaques and marmosets?

- They are relatively close to us phylogenetically, having diverged from the human lineage about 25 and 35 MYA, respectively (Fig. 1. Studying both Old-World (macaques) and New-World (marmosets) monkeys provides two evolutionary time points for comparison with humans, allowing testing for more complex patterns of evolutionary change than with a single comparison species (Wilson et al., 2013).
- Both species have complex, albeit fairly different, social behaviours that they regulate using very different sets of complex vocalizations well characterized acoustically (macaque: (Fukushima et al., 2015; Green, 1975; Hauser, 1991; Kalin et al., 1992); marmoset (Agamaite et al., 2015; DiMattina et al., 2006; Miller et al., 2010a; Pistorio et al., 2006; Turesson et al., 2016)).
- Both models are widely studied in neuroscience, in particularly in the auditory domain, providing large amounts of physiological, anatomical and neuroimaging data for reference (e.g., macaque: (Gil-da-Costa et al., 2006; Hackett, 2011; Kaas et al., 1999; Petkov et al., 2015; Poremba et al., 2004; Rauschecker and Tian, 2000; Recanzone, 2008; Tian et al., 2001); marmoset: (Bendor and Wang, 2005; Eliades and Wang, 2008, 2013; Newman et al., 2009; Nummela et al., 2017; Roy et al., 2016; Wang, 2000; Wang and Kadia, 2001; Wang et al., 1995). In particular, the marmoset is highly promising for the application of gene editing techniques in a primate model (Marx, 2016; Miller et al., 2016; Okano et al., 2016).
- Well-documented inter-species differences concerning their habitat (Brown, 2003), their vocal repertoire (Agamaite et al., 2015; Hauser and Marler, 1992; Owren et al., 1993; Rowell and Hinde, 1962) or their brain anatomy (de la Mothe et al., 2006; Nishimura et al., 2018) can provide additional knowledge on the constraints that shape vocal perception.
- Both species can be trained to perform awake fMRI, providing a unique bridge between the human fMRI and the monkey electrophysiological literature (Miller et al., 2010a; Silva, 2017; Vanduffel et al., 2014). Marmosets are a particularly promising model for awake fMRI as they can be trained in only a few weeks to tolerate an immobilizing cradle, considerably reducing the necessary training time compared to macaques, and eliminating the need for head-post surgery. Performing fMRI scanning of other non-human primates such as baboons or chimpanzees would be extremely valuable, but is not an option either for ethical reasons and because their strength makes them too dangerous for awake scanning.
- Macaques and marmosets are the only two species of non-human primates in whom human-like "voice patches" have been observed using awake fMRI (Petkov et al., 2008; Sadagopan et al., 2015).

We focus this survey on voice perception abilities that can be compared, i.e. *present in*, and *having adaptive significance for*, all three species. Two basic building blocks of voice perception (Belin et al., 2004), relatively well understood in humans, are examined: (1) the behavioural and neural sensitivity to CVs; and (2) the processing of speaker identity cues in CVs, allowing listeners to discriminate between individuals by voice.

2. Behavioural and neural sensitivity to CVs

2.1. Humans

Humans have remarkable abilities to extract information in voice – speech, but also identity, affect, personality, etc. (Belin et al., 2004, 2011; Kreiman, 1997; Kreiman and Sidtis, 2013)—perhaps because vocal sounds have such immense ecological relevance to us. Yet it is only quite recently that a behavioural advantage at voice detection has been experimentally demonstrated in human listeners. When presented with brief sounds and asked to decide whether they belong to a target category or not, listeners perform well even at very brief durations when the target category is Voice: 4 ms of sound are sufficient to yield above-chance performance at voice/non-voice discrimination, while at this very brief duration performance is at chance for other target categories (Suied et al., 2014). Moreover, when listeners are asked to detect a target sound category in a series of rapidly presented distracters performance is always better, across a range of experimental conditions, when the target category is Voice (Isnard, 2016).

Such behavioural sensitivity is paralleled by neural sensitivity to voice: secondary areas of human auditory cortex along the superior temporal gyrus (STG) and sulcus (STS) both anterior and posterior to primary auditory cortex contain temporal voice areas (TVAs) (Belin et al., 2000, 2002; Pernet et al., 2015; Von Kriegstein and Giraud, 2004) that show greater fMRI signal in response to vocal sounds—whether they contain speech or not—than to other categories of non-vocal sounds such as environmental sounds, amplitude-modulated noise, etc. (Agus et al., 2017; Belin et al., 2000; Von Kriegstein and Giraud, 2004) or to hetero-specific vocalizations (HVs) (Fecteau et al., 2004). The TVAs have been consistently observed by different groups including ours (Bestelmeyer et al., 2012, 2014; Bonte et al., 2013; Charest et al., 2013; Ethofer et al., 2009; Fecteau et al., 2004; Grandjean et al., 2005; Latinus et al., 2013; Leaver and Rauschecker, 2010; Lewis et al., 2009; Linden et al., 2011; Meyer et al., 2005; Pernet et al., 2015; Talkington et al., 2012; Von Kriegstein and Giraud, 2004). Although their exact anatomical location in the temporal lobe varies considerably across individuals, the TVAs are remarkably consistent within individuals in test-retest analysis (Pernet et al., 2015).

A cluster analysis of voice-sensitivity peaks in several hundred subjects suggests an organization in three "voice patches" along STG/STS bilaterally (TVAa, TVAm, TVAp; Fig. 2). That study also showed that the TVAs are essentially bilateral with no significant lateralization in activity overall, although more subjects (33%) showed significant right-sided than left-sided (13%) asymmetry in voice-sensitivity in the temporal lobe (Pernet et al., 2015). Voice processing also engages cerebral areas outside of auditory cortex, including several prefrontal areas (particularly in the inferior frontal gyrus bilaterally (Fecteau et al., 2005; Pernet et al., 2015).

The anatomo-functional organization of the TVAs remains poorly understood. Their causal link with voice processing has been established in a single study so far: transiently interfering with neuronal activity in the right TVAm via transcranial magnetic stimulation (TMS) interferes with performance at a voice detection task but not at a more general auditory task (Bestelmeyer et al., 2011).

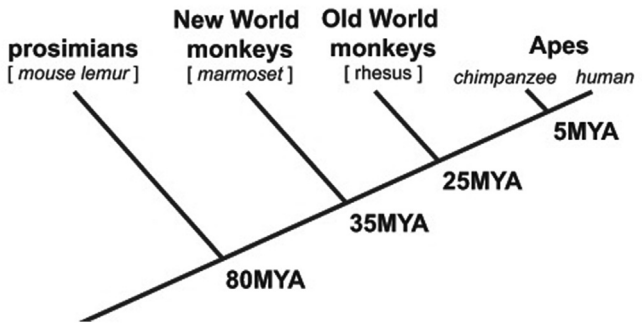

Fig. 1. Phylogenetic tree of primates. Cladogram showing the evolutionary divergence between humans and each of the main primate taxonomic groups with estimated time points of divergence (MYA, millions of years ago). Adapted withpermission from (Miller et al., 2016).

One issue stands out, that of nature vs. nurture: is voice-selectivity in the TVAs the result of evolutionary tuned innate mechanisms present at birth or does it reflect the extensive experience during development and adulthood with this ecologically crucial sound category? No convincing answer to that question has been provided so far, perhaps because long-term manipulation of human listeners' auditory environment is hard to perform.

2.2. Macaques

In the wild, macaques rely frequently on vocalizations to regulate and coordinate group activities using a rich call repertoire divided into 12–16 classes according to presumed context and motivational state (Hauser and Marler, 1992; Hauser, 1991; Rowell and Hinde, 1962). One classic series of studies found that Japanese macaques perform better than comparison species at discriminating between different CVs based on features (supposed to be) communication-relevant only for them (Petersen et al., 1978, 1984; Zoloth et al., 1979). Also, play back studies in the wild using the head turning paradigm report different patterns of ear preferences for CVs and hetero-specific vocalizations (Ghazanfar et al., 2001; Hauser and Andersson, 1994)—although with disputed results (Fitch and Fritz, 2006; Teufel et al., 2010). Thus, whether macaques show the same behavioural advantage as humans at detecting or discriminating CVs compared to other sounds remains essentially unknown.

The auditory cortex of macaques has been extensively investigated using multiple complementary techniques (cf. reviews in (Ghazanfar and Santos, 2004; Ghazanfar and Eliades, 2014; Hackett, 2011; Kaas et al., 1999; Rauschecker, 1998; Rauschecker and Scott, 2009; Romanski and Averbeck, 2009)). Electrophysiological recordings in awake animals show that neurons of belt and parabelt areas of secondary auditory cortex show strong sensitivity to CVs (Ghazanfar et al., 2008; Perrodin et al., 2011; Romanski and Averbeck, 2009; Tian et al., 2001) with latencies and selectivity increasing along the caudo-rostral direction towards the temporal pole (Fukushima et al., 2014; Kikuchi et al., 2010). The strong sensitivity of temporal lobe regions to CVs has been confirmed by the use of whole-brain metabolic imaging techniques (Gil-da-Costa et al., 2006; Poremba et al., 2004). Thanks to the development of macaque fMRI, whole-brain estimates of cerebral sensitivity to CV could be obtained using scanning protocols similar to those used in humans. Petkov et al. (2008) were the first to evidence a macaque voice area (Fig. 3a) with responses analogous to the human TVAs, i.e. areas with significantly stronger response to macaque CVs than to other categories of natural or control sounds. Specifically, at least two CV-preferring clusters were found: the first one was located bilaterally in the posterior auditory cortex, close to A1 region, whereas the second one was found in the high-hierarchical anterior portion of the right temporal lobe. Importantly, this anterior CV-preferring voice patch was still observed in anaesthetized monkeys, removing the possible effect of attention to sounds (Petkov et al., 2008). fMRI-guided electrophysiology in the anterior voice patch could further show that this area contains voice cells, i.e.

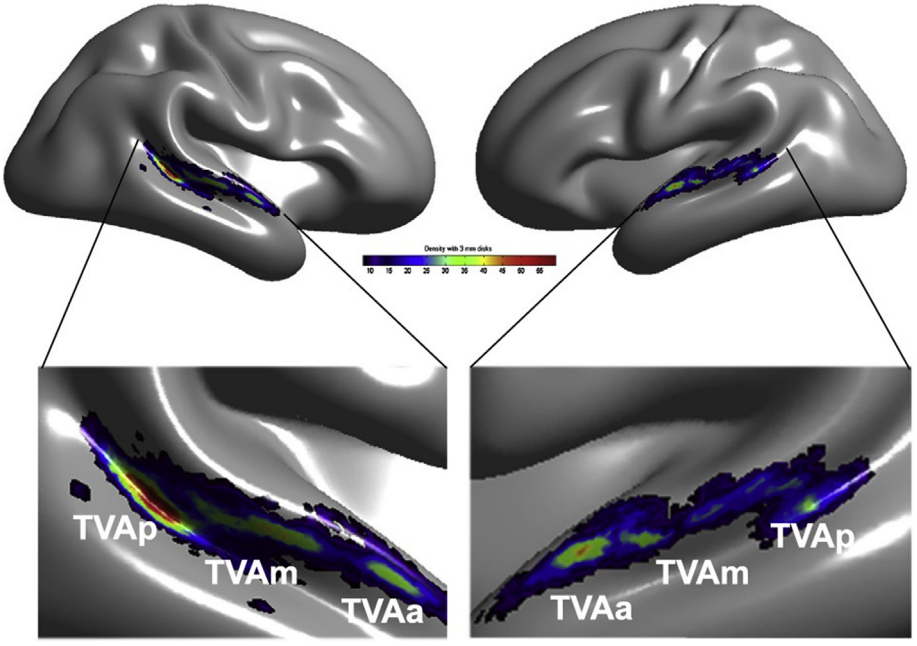

Fig. 2. The human Temporal Voice Areas (TVAs). The TVAs show greater fMRI response to vocal vs. non-vocal sounds; they are organized in three rostro-caudal “voice patches” along the STS and STG in the temporal lobe ofeach hemisphere. Reproduced from (Pernet et al., 2015).

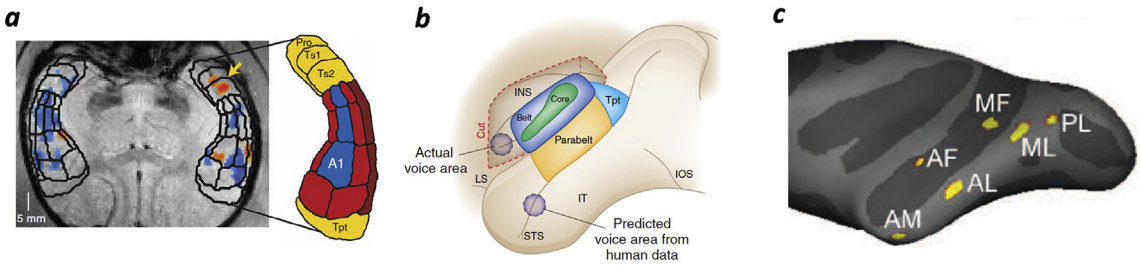

Fig. 3. Macaque voice and face areas. a. Macaque fMRI reveals (at least) one voice area (arrow) with strong preference for CVs in the anterior temporal lobe. Reproduced with permission from (Petkov et al., 2008). b. That macaque voice area does not seem to be located where expected based on direct human analogy. Reproduced with permission from (Ghazanfar, 2008). c. The macaque face patch system is increasingly well characterized both anatomically and functionally; here are shown the six face patches of the macaque left hemisphere. Reproduced with permission from (Freiwald and Tsao, 2010).

individual neurons showing voice-selectivity (Perrodin et al., 2011), analogous to results obtained in face patches (Tsao et al., 2006).

Joly and colleagues (2012) performed a pioneering comparative study in which human and macaque subjects were scanned while exposed to the same stimuli including macaque and human vocalizations (Joly et al., 2012b). Areas along STG close to primary auditory cortex in macaques showed greater response to CVs compared to acoustical control sounds, but no compared to human vocalizations (HVs). Inter-species comparison suggested that CVs recruited distinct regions, mainly in the STG/lateral sulcus in macaques and along the STS in humans. A more recent study also reported CV selectivity in the STG compared to environmental and spectro-temporally controlled sounds (Ortiz-Rios et al., 2015). Both middle and anterior portions of the STG were highlighted, as in Petkov et al. (2008). CV-preferring clusters were also identified in the ventral lateral prefrontal cortex, a region that seems implicated in call type categorization according to previous electrophysiological reports (Averbeck and Romanski, 2006; Gifford et al., 2003; Romanski et al., 2005).

Evidence of voice areas in the auditory cortex of macaques is suggestive of gradual evolution rather than abrupt changes of the neural structures involved in vocal communication (Ghazanfar, 2008). However, as could be expected based on only a handful of reports by different groups using different protocols there are discrepancies in the observed pattern of results, emphasizing the need for replication and extension of these seminal studies. Interestingly, current evidence seems to suggest that the position of the anterior macaque voice area is quite different from what would be expected from human data (Ghazanfar, 2008; Ghazanfar and Eliades, 2014) (Fig. 3b) highlighting the need for more precise comparisons using complementary measures such as anatomical connectivity (Rilling, 2014a). The existence of other CV-sensitive areas shown but not emphasized in Petkov et al. (2008)'s results and also observed in the other studies (Joly et al., 2012a, 2012b; Ortiz-Rios et al., 2015) suggests that there could be several voice patches in the macaque brain as in the human brain (Fig. 2), potentially organized in a network of interconnected voice patches comparable to the face patches network of visual cortex (Fig. 3c) (Chang and Tsao, 2017; Freiwald and Tsao, 2010; Freiwald et al., 2009; Meyers et al., 2015; Tsao et al., 2006).

2.3. Marmosets

Marmosets are a highly vocal species, engaging in nearly constant vocal communication even in captivity (cf. reviews in (Eliades and Miller, 2017; Miller et al., 2016)). Their vocalization repertoire, well characterized acoustically (Agamaite et al., 2015; Miller et al., 2010a; Pistorio et al., 2006), includes several types of calls produced depending on social and ecological context, including "twitters" (series of short, rapid frequency modulated sounds, cf. Fig. 7b) and "trills" (with sinusoidal frequency modulation) both apparently mediating interactions in close proximity although their explicit function remains unclear. They also produce "phees" (slow frequency-modulated whistle-like tones) to maintain long-distance contact with other group members, sometimes in dialogs of alternating calls (antiphonal calling) by different callers (Miller et al., 2010a). Although there is clear observational evidence that marmosets detect and extract information from CVs, whether they have a particular sensitivity to CVs compared to other sound categories, and whether they can be trained to discriminate CVs from non-CVs, is not established.

The auditory cortex of marmosets is thought to be organized similarly to that of the macaque with core, belt and parabelt areas with increasingly complex receptive fields. A series of elegant neurophysiological studies with recordings performed in freely moving and interacting individuals has characterized the response properties of neurons in auditory core areas, where neuronal populations show strong sensitivity to CVs (Nagarajan et al., 2002; Wang and Kadia, 2001; Wang et al., 1995) reflecting in particular the activity of harmonic template neurons (Feng and Wang, 2017). However, whole-brain measures of neuronal activity using cFOS expression quantification (Miller et al., 2010b) or electrophysiological recordings outside of temporal lobe suggest that the perception of CVs engages a number of cerebral areas beyond core auditory areas, including areas of prefrontal cortex (Nummela et al., 2017), as suggested by early studies in a close cousin the squirrel monkey (Glass and Wollberg, 1983; Winter and Funkenstein, 1973; Wollberg and Newman, 1972).

Recent developments in marmoset MRI imaging hold much promise (Belcher et al., 2013; Hung et al., 2015a, 2015b; Papoti et al., 2013, 2017), particularly as its small size is compatible with high-field (7T) rodent MRI allowing for higher signal and spatial resolution to compensate for their small brain size. Remarkably, a recent fMRI study in anaesthetized marmosets has revealed a gradient of sensitivity to vocalizations along a caudal-ventral axis (Sadagopan et al., 2015), with areas of high selectivity to CVs, or voice patches, in the most anterior parts of temporal lobe bilaterally (Fig. 4a). This recent finding, that needs to be replicated in awake, behaving animals, suggests that the processing of CVs in the marmoset brain could be performed as in humans and macaques by an array of interconnected voice patches similar to that observed for face processing (Fig. 4b).

Thus, humans show particular neural sensitivity to sounds of voice with voice-selective "temporal voice areas (TVAs)" organized in three "voice patches" bilaterally. Initial evidence suggests the existence of CV-selective voice patches potentially homologous to the human TVAs in both macaques and marmosets. Our hypothesis outlined below is that these findings reflect the existence of a

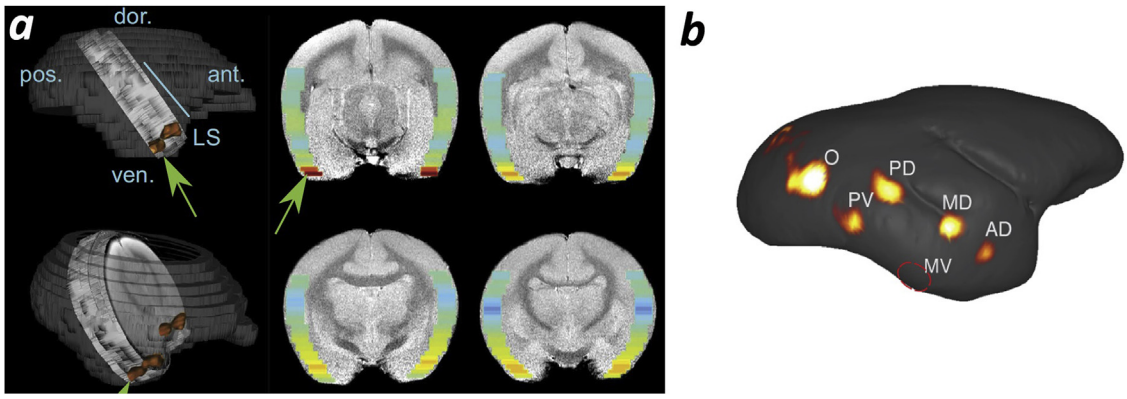

Fig. 4. Marmoset voice and face areas. a. High-field fMRI in anaesthetized marmosets reveals a gradient of selectivity to CVs with focal voice patches inthe anterior temporal lobe bilaterally. Reproduced with permission from (Sadagopan et al., 2015). b. Face-selective patches measured obtained by the contrast faces > objects in awake marmoset fMRI are highly analogous to those of the macaque (Fig. 3c). Figure courtesy of Afonso Silva.

network of interconnected voice patches involved in detecting and extracting information from CVs and that this voice patch system, similar to the face patch system of visual cortex, and potentially *conserved in primates*. The role of auditory experience in shaping CV selectivity in that network is not understood yet.

3. Processing of speaker/caller identity cues

3.1. Humans

Human listeners possess to variable degrees the ability to discriminate unfamiliar identities based on voice and the ability to recognize familiar identities in novel utterances (reviews in (Blank et al., 2014; Kreiman and Sidtis, 2013; Schweinberger et al., 2014)), two dissociable abilities (Van Lancker and Kreiman, 1987). Van Lancker et al. (1988) observed for the first time impaired voice recognition occurring after stroke, a deficit that took the name of "phonagnosia" (Van Lancker et al., 1988), to mirror prosopagnosia, the deficit occurring for face recognition. This deficit can indeed be acquired after stroke (acquired phonagnosia) but it can also be present from birth, notwithstanding intact brain structures and normal auditory abilities (developmental phonagnosia). Since the discovery of the first case of developmental phonagnosia (Garrido et al., 2009), other three cases have been documented in literature showing either impaired recognition of famous voices or an impairment in unfamiliar voice recognition (Roswandowitz et al., 2014; Xu et al., 2015). As put forward by Russell et al. (2009) in the domain of face perception, subjects affected by developmental phonagnosia could be thought of as extreme cases of the broad distribution of individual differences in voice recognition abilities, while the extreme cases at the opposite tail of the distribution can be referred to as "super-recognizers" (Russell et al., 2009). Recently, we demonstrated that the scores obtained by a big cohort of subjects (1000) at the Glasgow Voice Memory Test, a 5−minutes test assessing unfamiliar voice recognition, indeed spanned from significantly poor performances (potential developmental phonagnosia) to perfect voice recognition (super-recognizers) (Aglieri et al., 2017). These behavioural individual differences in voice recognition could have their neural correlates in the considerable inter-individual variability observed in voice-elicited BOLD responses (Pernet et al., 2015). Nonetheless, the cognitive and neural mechanisms behind inter-individual variability in voice recognition remain, to date, poorly understood.

The discrimination of unfamiliar speakers appears to obey the voice space metaphor, inspired from the face recognition literature (Chang and Tsao, 2017; Freiwald et al., 2009; Valentine, 1991): each voice can be viewed as a point in a multidimensional space with dimensions corresponding to auditory features used to discriminate speakers; voices close to one another in that space are hard to discriminate from one another, while voices far apart are easily discriminable (Baumann and Belin, 2010; Latinus and Belin, 2011; Latinus et al., 2013). Using multidimensional scaling analyses of identity discrimination performance for many speaker pairs (Baumann and Belin, 2010) we showed that the two main dimensions of the voice space in human listeners are f0 (fundamental frequency, reflecting the rate of vocal fold oscillation) and formant dispersion (average frequency difference between formant, or vocal tract resonances, reflecting vocal tract size (Fitch, 2000; González, 2004)); harmonic-to-noise ratio (HNR), reflecting voice irregularities provides a third important dimension (Latinus et al., 2013). Notably, voice perception in that space follows norm-based coding: voices closer in voice space to a voice prototype (well approximated by the morphing-generated average of many speakers of the same gender) are perceived as less distinctive than voices less acoustically similar (farther away in voice space) to the prototype (Latinus et al., 2013). Human listeners are particularly accurate at voice gender recognition (Kreiman, 1997; Mullennix et al., 1995), using a combination of f0 and formant cues (Pernet and Belin, 2012)—reflecting the fact that both source and filter aspects of human voice production are strongly sexually dimorphic (Titze, 1989). Indeed norm-based coding is based on two male and female voice prototypes (Latinus et al., 2013).

The cerebral processing of speaker identity involves both temporal lobe and prefrontal regions with strong right-hemispheric lateralization (Andics et al., 2010, 2013; Belin and Zatorre, 2003; Bonte et al., 2014; Formisano et al., 2008; Kriegstein and Giraud, 2004; Nakamura et al., 2001). The most anterior voice-sensitive region of the right temporal lobe (right TVAa) shows adaptation to speaker identity, i.e., smaller response to syllables spoken by a single speaker than to syllables spoken by multiple speakers (Belin and Zatorre, 2003) and is more active when listeners focus attention on speaker identity as opposed to sentence meaning (Kriegstein and Giraud, 2004). Studies using multi-voxel pattern analysis (MVPA) (Haxby et al., 2014) beautifully confirm this dissociation: whereas voxels most informative for classifying vowels are distributed bilaterally, those most informative for classifying speaker identity are mostly distributed along right STG/STS particularly its more anterior part (Bonte et al., 2014; Formisano et al., 2008). Unfamiliar voices are coded in the TVAs using norm-based coding, confirming behavioural evidence: voices

acoustically close to their (own-gender) prototype elicit smaller TVA activity than more distinctive, acoustically dissimilar voices (Fig. 5b) (Latinus et al., 2013). (Note that short-term adaptation has been ruled out as an explanation for this result (cf. (Kahn and Aguirre, 2012)) but that the role of *long*-term experience remains unclear in shaping the voice prototypes.) Inferior prefrontal regions are involved in the learning of new voice identities (Latinus et al., 2011; Zäske et al., 2017), also with strong right-hemispheric lateralization, and use norm-based coding for representing familiar identities (Andics et al., 2013).

3.2. Macaques

There is clear behavioural evidence that macaques are able to use identity information in CVs (Gouzoules et al., 1984; Hauser, 1991, 1996; Petersen et al., 1978; Zoloth et al., 1979). In the wild, female macaques respond appropriately to playbacks of screams from their immature offspring (Gouzoules et al., 1984); they respond faster and longer to coos from by matrilineal relatives, and show rebound of habituation for coos produced by different relatives, demonstrating an ability for vocal recognition of both individual and kin (Rendall et al., 1996). Interestingly, macaques appear to also use formant frequency information (related to vocal tract and body size in macaques (Fitch, 1997) and the main acoustical cue to human phonemes): not only do macaques spontaneously perceive formant frequency changes in playback trials (Fitch and Fritz, 2006), but they also associate these changes to differences in perceived body size (Ghazanfar et al., 2007) as humans do. It is unclear, however, whether macaques represent different callers in a "macaque voice space" and what would be the underlying acoustical dimensions. Although macaques show moderate sexual dimorphism in body size, males being on average slightly larger and heavier, it is not even clear whether macaques can recognize caller gender.

The cerebral bases of caller identity processing in macaques have only begun to be investigated (reviewed in (Perrodin et al., 2015)). The anterior voice area observed in macaques shows the same speaker adaptation response observed in humans in the analogous area of right anterior temporal lobe (Belin and Zatorre, 2003): greater response to CVs from different individuals than to CVs from a single individual (Petkov et al., 2008) (Fig. 6a). Some of the voice cells in that region also show some degree of caller selectivity (Fig. 6b), differentiating between individuals more than call type (Perrodin et al., 2014)—a finding reminiscent of face processing results (Freiwald and Tsao, 2010).

3.3. Marmosets

Acoustical analyses indicate that some marmoset vocalizations such as antiphonal phees and thrills can potentially convey important identity information, being quite variable between individuals despite a fixed structure (Agamaite et al., 2015; Miller et al., 2010a). Indeed, there is experimental evidence of voice identity discrimination in marmosets. Using an ingenious automated playback technique exploiting the antiphonal calling behaviour of marmosets (the "Virtual Monkey" approach (Miller and Wren Thomas, 2012; Toarmino et al., 2017)), Cory Miller and Wren Thomas, 2012 showed that changes in the identity of synthetic phees were followed by changes in the frequency and latency of antiphonal calling by the subject, demonstrating identity discrimination by voice alone. However to our knowledge there is no experimental evidence yet relevant to the neural coding of caller identity in marmosets.

In summary, human listeners appear to represent unfamiliar speaker identity using norm-based coding relative to gender-specific voice prototypes in a voice space with main dimensions related to f0 and formant frequencies. This involves neuronal populations in the TVAs and inferior prefrontal regions with strong right-hemispheric lateralization. The role of auditory experience in shaping the prototypes remains unclear. The scarce data available (in only two animals) also suggests a role of the macaque right anterior voice patch in coding caller identity; to our knowledge, no relevant evidence is available yet in marmosets.

4. A primate "voice patch system" for cerebral processing of voice information

The evidence reviewed above naturally leads to the notion of a "voice patch" system in the primate temporal lobe dedicated to processing information in CVs. Such a voice patch system could be the auditory counterpart of the "face patch" system of inferotemporal cortex (Figs. 3c and 4b). Studies in humans, macaques (Fig. 3c) and, more recently, marmosets (Fig. 4b) together demonstrate the existence of a system of discrete, interconnected face-sensitive areas containing "face cells" and supporting a series of increasingly abstract (identity-invariant) face representations (Chang and Tsao, 2017; Freiwald and Tsao, 2010; Freiwald et al., 2009; Meyers et al., 2015; Tsao et al., 2006). Moreover the face patch system appears largely conserved in primates such that the macaque face patch system is widely considered as a simpler, less variable model of the human face areas (indeed the macaque face patch system is often probed with human faces!)

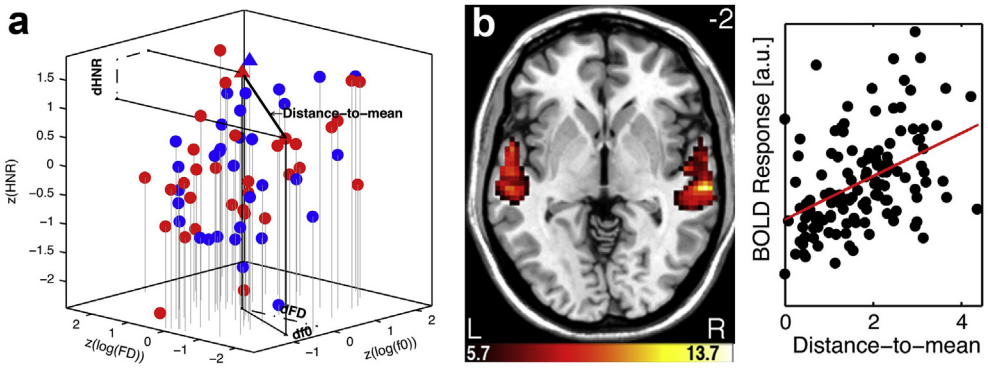

Fig. 5. Norm-based coding of voice identity in the human TVAs. a. Voices aspoints in a 3D "Voice space" with dimensions reflecting: f0, formant dispersion, harmonics-to-noise ratio. b. The Euclidian distance between a voice and its same-gender prototype ("distance-to-mean") is a strong predictor of the voice's evoked neural activity in right TVAm. Reproduced from (Latinus et al.,2013).

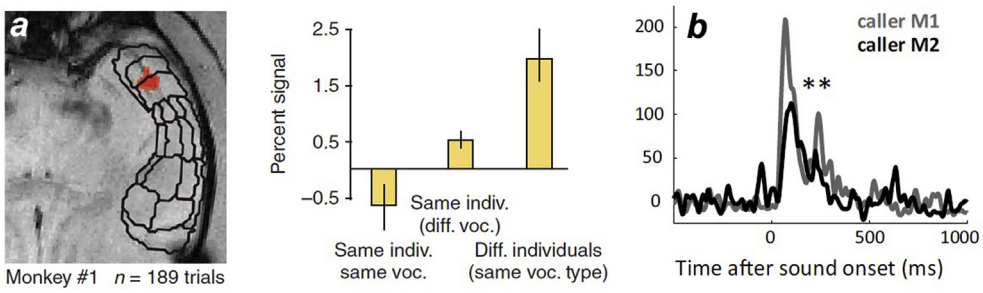

Fig. 6. Identity processing in the macaque brain. a. The anterior voice patch shows adaptation to speaker identity (Petkov et al., 2008) exactly as in humans (Belin and Zatorre, 2003). Reproduced with permission from (Petkov et al., 2008). b. Some "voice cells" in that region are selective to caller identity independently of the vocalization. Reproduced with permission from (Perrodin et al., 2014).

In humans the face-voice analogy is indeed powerful at explaining and predicting several properties of the voice-processing system: category-selective cortex, norm-based coding, causal link with perception, etc. (reviewed in (Yovel and Belin, 2013)). This appears a parsimonious principle of organization as the computational problems to be solved (detection, invariance, …) of a similar nature in both modalities, and similar processing architectures clearly facilitate integration from the auditory and visual modalities in natural polymodal environments (Belin et al., 2004; Campanella and Belin, 2007).

The face-voice analogy extended to other primates suggests that the voice patch system could also be organized as a series of discrete, interconnected voice patches supporting increasingly abstract voice representations from template matching to speaker-invariant representations. As in the face perception domain, these functional differences could be reflected in the spatial organization of the patches, with a gradient of selectivity and abstraction from core areas to the temporal pole region. A key, open question is whether such a primate voice patch system would be as conserved as the face patch systems apparently is, or whether the emergence of speech and language in hominins has dramatically altered the mechanisms of CV sensitivity and speaker identity processing such that its organization would be very different in humans vs. macaques and marmosets. That question awaits testing in particular by adapting to macaques and marmosets experimental paradigms derived from human research to directly and quantitatively compare the underlying perceptual and neural mechanisms. Statistically testing for differences in behavioural and neural indices of voice perception across the three species have the potential to provide rigorous tests of the hypothesis and enable reconstruction of the evolution of voice perception in the form of inter-specific distances.

5. Challenges in probing the primate voice patch system

To fill the wide gap in our understanding of the differences in the neurocognitive bases of CV sensitivity and identity processing in humans and monkeys, as well as in the methodology to assess these differences, several challenges arise. A *first challenge* lies in the notorious difficulty of training monkeys, particularly macaques, to perform auditory perceptual tasks. This has hindered more widespread efforts than the handful of playback experiments in the wild or behavioural tests in limited samples of laboratory animals that current evidence is based upon. Such difficulty could be related to species-specific differences in auditory attention (Rinne et al., 2017) or long-term memory (Fritz et al., 2005), but also to inefficient auditory training paradigms. One exciting opportunity is provided by the large-scale behavioural testing developed by Joël Fagot (CNRS, Marseille) for baboons with outstanding results (Fagot and Paleressompoulle, 2009; Fagot and Bonte, 2010; Grainger et al., 2012). This method, that relies on ad-lib access to testing systems by monkeys living in a large social group, has proven highly successful in allowing collection of over a million trials in a few weeks by a group of baboons (Fagot and Bonte, 2010), and has been shown to work in macaques as well (Fagot and Paleressompoulle, 2009). Such methodology would be extremely valuable for comparing voice perception behaviour across primates; psychometric response functions obtained in the three species would allow direct quantitative comparison and estimation of inter-specific distances. The power afforded by the potentially very large number of trials could even be exploited in reverse-correlation experiments.

A *second challenge* lies in the comparison of measures of neural activity in monkeys and humans—so far largely based on electrophysiological recordings in 1–3 monkeys vs. whole-brain neuroimaging in 10–30 humans. fMRI emerges as the technique of choice for direct comparison of measures of neuronal activity in groups of awake, behaving subjects in all three species. Macaque fMRI is now well established as a method of choice for bridging human fMRI and macaque electrophysiology and is being used by an increasing number of groups (Vanduffel et al., 2014) including our own; excellent recent developments (Hung et al., 2015a; Silva, 2017; Toarmino et al., 2017) in awake marmoset fMRI suggest it is now possible to measure neural activity using the same technique in awake subjects across the three species.

A *third challenge* lies in understanding the role of long-term auditory experience in shaping behavioural and neural sensitivity to CVs and the neural coding of speaker identity. There is evidence that experience during the first months of life significantly alters preference for voice over other sounds (Vouloumanos et al., 2010) as well as the coding of own-language phonemes (Kuhl, 1994). But how does exposure to sound over the long-term affect TVA selectivity in the adult brain? Would a similar type of selectivity emerge for other, behaviourally relevant categories after intense exposure (cf. (Gauthier et al., 2000), for similar question in the visual domain)? Are the voice prototypes fixed, genetically-encoded templates or are they the weighted average of all voices heard in one's lifetime—or the past few months? Animal models offer an opportunity for subject-specific, long-term manipulation of the auditory stimulation while preserving high standards of welfare, potentially providing unique insight into the experience-dependence of voice-patch selectivity and coding mechanisms in monkeys, possibly analogous in humans.

Finally, a *fourth challenge* lies in understanding which features drive neuronal responses in the acoustically complex and variable CVs. One strategy that has been used successfully in particular in the domain of avian vocal communication (e.g. (Gentner and Margoliash, 2003),) is to use artificial, synthetic models of the CVs that can be manipulated in specific, rigorously controlled ways

and their effect on behavioural and neural responses monitored. This approach has already been used independently in humans in particular by my group (Bestelmeyer et al., 2010; Charest et al., 2013; Latinus and Belin, 2010, 2011, 2012; Latinus et al., 2011), in macaques (Chakladar et al., 2008; Ghazanfar et al., 2007) and in marmosets in whom "virtual vocalization" models have been established for several call types and identities (DiMattina and wang, 2006) as well as automated "virtual monkey" software developed for eliciting antiphonal calling and testing perceptual differences (Miller and Wren Thomas, 2012; Toarmino et al., 2017).

6. Conclusion

We have reviewed evidence available in humans, macaques and marmosets on the perceptual and neural mechanisms involved in detecting CVs and processing identity cues in CVs. It is still too fragmentary for the in-depth comparisons in behavioural, anatomical and functional mechanisms required by the comparative approach for a detailed reconstruction of the recent evolution of voice perception and precise characterization of the vocal brain of our common ancestor. Yet current evidence is compatible with the notion of a network of discrete, interconnected cortical "voice patches" in the primate brain carrying out different operations in a complex functional architecture for voice information processing. The degree to which this primate voice patch system is conserved in humans, and to which it has been modified by the emergence of speech and language, remains to be investigated using a combination of techniques such as fMRI and voice morphing for comparable experimental protocols in the three species.

Acknowledgements

Supported by grants from the French Fondation pour la Recherche Médicale (AJE201214) and Agence Nationale de la Recherche to PB, and by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heares.2018.04.010.

References

Agamaite, J.A., Chang, C.J., Osmanski, M.S., Wang, X., 2015. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). J. Acoust. Soc. Am. 138, 2906–2928.
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. Behav. Res. Methods 49, 97–110.
Agus, T.R., Paquette, S., Suied, C., Pressnitzer, D., Belin, P., 2017. Voice selectivity in the temporal voice area despite matched low-level acoustic cues. Sci. Rep. 7, 11526.
Andics, A., McQueen, J.M., Petersson, K.M., 2013. Mean-based neural coding of voices. NEUROIMAGE 79, 351–360.
Andics, A., McQueen, J.M., Petersson, K.M., Gál, V., Rudas, G., Vidnyánszky, Z., 2010. Neural mechanisms for voice recognition. NEUROIMAGE 52, 1528–1540.
Averbeck, B.B., Romanski, L.M., 2006. Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. J. Neurosci. 26, 11023–11033.
Baumann, O., Belin, P., 2010. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. Psychol. Research-Psychologische Forsch. 74, 110–120.
Belcher, A.M., Yen, C.C., Stepp, H., Gu, H., Lu, H., Yang, Y., Silva, A.C., Stein, E.A., 2013. Large-scale brain networks in the awake, truly resting marmoset monkey. J. Neurosci. 33, 16796–16804.
Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105–2109.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cognitive Brain Res. 13, 17–26.
Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.
Belin, P., Bestelmeyer, P.E., Latinus, M., Watson, R., 2011. Understanding voice perception. Br. J. Psychol. 102, 711–725.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165.
Bestelmeyer, P., Belin, P., Grosbras, M.H., 2011. Right temporal TMS impairs voice detection. Curr. Biol. 21, R838–R839.
Bestelmeyer, P., Rouger, J., DeBruine, L.M., Belin, P., 2010. Auditory adaptation in vocal affect perception. Cognition 117, 217–223.
Bestelmeyer, P.E., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014. Adaptation to vocal expressions reveals multistep perception of auditory emotion. J. Neurosci. 34, 8098–8105.
Bestelmeyer, P.E., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., Belin, P., 2012. Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. Cereb. Cortex 22, 1263–1270.
Blank, H., Wieland, N., von Kriegstein, K., 2014. Person recognition and the brain: merging evidence from patients and healthy individuals. Neurosci. Biobehav Rev. 47, 717–734.
Boe, L.J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T.R., Becker, Y., Rey, A., Fagot, J., 2017. Evidence of a vocalic proto-system in the baboon (Papio papio) suggests pre-hominin speech precursors. PLoS One 12, e0169321.
Bonte, M., Hausfeld, L., Scharke, W., Valente, G., Formisano, E., 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J. Neurosci. 34, 4548–4557.
Bonte, M., Frost, M.A., Rutten, S., Ley, A., Formisano, E., Goebel, R., 2013. Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. NEUROIMAGE 83, 739–750.
Brown, C.H., 2003. Ecological and physiological constraints for primate vocal communication. In: Ghazanfar, A.A. (Ed.), Primate Audition: Ethology and Neurobiology. CRC Press, pp. 127–150.
Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. Trends Cogn. Sci. 11, 535–543.
Chakladar, S., Logothetis, N.K., Petkov, C.I., 2008. Morphing rhesus monkey vocalizations. J. Neurosci. Methods 170, 45–55.
Chang, L., Tsao, D.Y., 2017. The code for facial identity in the primate brain. Cell 169, 1013–1028 e14.
Charest, I., Pernet, C., Latinus, M., Crabbe, F., Belin, P., 2013. Cerebral processing of voice gender studied using a continuous carryover FMRI design. Cereb. Cortex 23, 958–966.
de la Mothe, L.A., Blumell, S., Kajikawa, Y., Hackett, T.A., 2006. Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J. Comp. Neurol. 496, 27–71.
DiMattina, C., Wang, X., 2006. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. J Neurophysiol 95, 1244–1262. https://doi.org/10.1152/jn.00818.2005.
DiMattina, C., Wang, X., 2006. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. J. Neurophysiol. 95, 1244–1262.
Eliades, S.J., Wang, X., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–1106.
Eliades, S.J., Wang, X., 2013. Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. J. Neurophysiol. 109, 1638–1657.
Eliades, S.J., Miller, C.T., 2017. Marmoset vocal communication: behavior and neurobiology. Dev. Neurobiol. 77, 286–299.
Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. Curr. Biol. 19, 1028–1033.
Fagot, J., Paleressompoulle, D., 2009. Automatic testing of cognitive performance in baboons maintained in social groups. Behav. Res. Methods 41, 396–404.
Fagot, J., Bonte, E., 2010. Automated testing of cognitive performance in monkeys: use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). Behav. Res. Methods 42, 507–516.
Fecteau, S., Armony, J.L., Joanette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. NEUROIMAGE 23, 840–848.
Fecteau, S., Armony, J.L., Joanette, Y., Belin, P., 2005. Sensitivity to voice in human prefrontal cortex. J. Neurophysiol. 94, 2251–2254.
Feng, L., Wang, X., 2017. Harmonic template neurons in primate auditory cortex underlying complex sound processing. Proc. Natl. Acad. Sci. U. S. A. 114, E840–E848.
Fitch, W.T., 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. J. Acoust. Soc. Am. 102, 1213–1222.
Fitch, W.T., 2000. The evolution of speech: a comparative review. Trends Cogn. Sci. 4, 258–267.
Fitch, W.T., Fritz, J.B., 2006. Rhesus macaques spontaneously perceive formants in conspecific vocalizations. J. Acoust. Soc. Am. 120, 2132–2141.
Fitch, W.T., de Boer, B., Mathur, N., Ghazanfar, A.A., 2016. Monkey vocal tracts are speech-ready. Sci. Adv. 2, e1600723.
Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "What"? Brain-Based decoding of human voice and speech. Science 322, 970–973.
Freiwald, W.A., Tsao, D.Y., 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330, 845–851.

Freiwald, W.A., Tsao, D.Y., Livingstone, M.S., 2009. A face feature space in the macaque temporal lobe. Nat. Neurosci. 12, 1187–1196.
Fritz, J., Mishkin, M., Saunders, R.C., 2005. In search of an auditory engram. PNAS 102, 9359–9364.
Fukushima, M., Saunders, R.C., Leopold, D.A., Mishkin, M., Averbeck, B.B., 2014. Differential coding of conspecific vocalizations in the ventral auditory cortical stream. J. Neurosci. official J. Soc. Neurosci. 34, 4665–4676.
Fukushima, M., Doyle, A.M., Mullarkey, M.P., Mishkin, M., Averbeck, B.B., 2015. Distributed acoustic cues for caller identity in macaque vocalization. R. Soc. open Sci. 2, 150432.
Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., Duchaine, B., 2009. Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 47, 123–131.
Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191–197.
Gentner, T.Q., Margoliash, D., 2003. Neuronal populations and single cells representing learned auditory objects. Nature 424, 669–674.
Ghazanfar, A.A., 2008. Language evolution: neural differences that make a difference. Nat. Neurosci. 11, 382–384.
Ghazanfar, A.A., Santos, L.R., 2004. Primate brains in the wild: the sensory bases for social interactions. Nat. Rev. Neurosci. 5, 603–616.
Ghazanfar, A.A., Rendall, D., 2008. Evolution of human vocal production. Curr. Biol. 18, R457–R460.
Ghazanfar, A.A., Eliades, S.J., 2014. The neurobiology of primate vocal communication. Curr. Opin. Neurobiol. 28, 128–135.
Ghazanfar, A.A., Smith-Rohrberg, D., Hauser, M.D., 2001. The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. Brain Behav. Evol. 58, 163–172.
Ghazanfar, A.A., Chandrasekaran, C., Logothetis, N.K., 2008. Interactions between the Superior Tempoal Sulcus and auditory cortex mediate dynamica face/voice integration in rhesus monkeys. J. Neurosci. 28, 4457–4469.
Ghazanfar, A.A., Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., Logothetis, N.K., 2007. Vocal-tract resonances as indexical cues in rhesus monkeys. Curr. Biol. 17, 425–430.
Gifford, G.W.I., Hauser, M.D., Cohen, Y.E., 2003. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. Brain, Behav. Evol. 61, 213–224.
Gil-da-Costa, R., Martin, A., Lopes, M.A., Munoz, M., Fritz, J.B., Braun, A.R., 2006. Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. Nat. Neurosci. 9, 1064–1070.
Glass, I., Wollberg, Z., 1983. Responses of cells in the auditory cortex of awake squirrel monkeys to normal and reversed species-specific vocalizations. Hear Res. 9, 27–33.
González, J., 2004. Formant frequencies and body size of speaker: a weak relationship in adult humans. J. Phon. 32, 277–287.
Gouzoules, S., Gouzoules, H., Marler, P., 1984. Rhesus monkey (Macaca Mulatta) screams: representational signalling in the recruitment of agonistic aid. Anim. Behav. 32, 182–193.
Grainger, J., Dufau, S., Montant, M., Ziegler, J.C., Fagot, J., 2012. Orthographic processing in baboons (Papio papio). Science 336, 245–248.
Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8, 145–146.
Green, S., 1975. Variation of vocal pattern with social situation in the Japanese monkey (Macaca fuscata): a field study. In: Rosenblum, L.A. (Ed.), Primate Behavior, vol. 4. Academic Press, New-York, pp. 1–102.
Hackett, T.A., 2011. Information flow in the auditory cortical network. Hear Res. 271, 133–146.
Hauser, M., 1996. The Evolution of Communication. MIT Press, Cambridge, Massachusetts London, England.
Hauser, M., Marler, P., 1992. Food-associated calls in rhesus mcaques (Macaca mulatta): I. Socioecological factors. Behav. Ecol. 4, 194–212.
Hauser, M.D., 1991. Sources of acoustic variation in rhesus macaque (Macaca mulatta) vocalizations. Ethology 89, 29–46.
Hauser, M.D., Andersson, K., 1994. Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: field experiments. Proc. Natl. Acad. Sci. U. S. A. 91, 3946–3948.
Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456.
Hung, C.C., Yen, C.C., Ciuchta, J.L., Papoti, D., Bock, N.A., Leopold, D.A., Silva, A.C., 2015a. Functional MRI of visual responses in the awake, behaving marmoset. NEUROIMAGE 120, 1–11.
Hung, C.C., Yen, C.C., Ciuchta, J.L., Papoti, D., Bock, N.A., Leopold, D.A., Silva, A.C., 2015b. Functional mapping of face-selective regions in the extrastriate visual cortex of the marmoset. J. Neurosci. 35, 1160–1172.
Isnard, V., 2016. L'efficacité du Système Auditif Humain Pour la Reconnaissance de Sons Naturels. Unpublished PhD thesis. Université Paris 6 Pierre et Marie Curie.
Joly, O., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012a. Interhemispheric differences in auditory processing revealed by FMRI in awake rhesus monkeys. Cereb. Cortex 22, 838–853.
Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012b. Processing of vocalizations in humans and monkeys: a comparative fMRI study. NEUROIMAGE 62, 1376–1389.
Kaas, J.H., Hackett, T.A., Tramo, M.J., 1999. Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. 9, 154–170.
Kahn, D.A., Aguirre, G.K., 2012. Confounding of norm-based and adaptation effects in brain responses. NEUROIMAGE 60, 2294–2299.
Kalin, N.H., Shelton, S.E., Snowdon, C.T., 1992. Affiliative vocalizations in infant rhesus macaques (Macaca mulatta). J. Comp. Psychol. 106, 254–261.
Kikuchi, Y., Horwitz, B., Mishkin, M., 2010. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. J. Neurosci. 30, 13021–13030.
Kreiman, J., 1997. Listening to voices: theory and practice in voice perception research. In: Johnson, K., Mullenix, J. (Eds.), Talker Variability in Speech Research. Academic Press, New-York, pp. 85–108.
Kreiman, J., Sidtis, D., 2013. Foundations of Voice Studies. An Interdisciplinary Approach to Voice Production and Perception. Wiley-Blackwell.
Kriegstein, K.V., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NEUROIMAGE 22, 948–955.
Kuhl, P.K., 1994. Learning and representation in speech and language. Curr. Opin. Neurobiol. 4, 812–822.
Latinus, M., Belin, P., 2010. Auditory Aftereffects Reveal Prototype-based Coding of Voice Identity Cognitive Neuroscience Meeting, Montreal.
Latinus, M., Belin, P., 2011. Anti-voice adaptation suggests prototype-based coding of voice identity. Front. Psychol. 2, 175.
Latinus, M., Belin, P., 2012. Perceptual auditory aftereffects on voice identity using brief vowel stimuli. PLoS One 7, e41384.
Latinus, M., Crabbe, F., Belin, P., 2011. Learning-induced changes in the cerebral processing of voice identity. Cereb. Cortex 21, 2820–2828.
Latinus, M., McAleer, P., Bestelmeyer, P.E., Belin, P., 2013. Norm-based coding of voice identity in human auditory cortex. Curr. Biol. 23, 1075–1080.
Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. 30, 7604–7612.
Lewis, J.W., Talkington, W.J., Walker, N.A., Spirou, G.A., Jajosky, A., Frum, C., Brefczynski-Lewis, J.A., 2009. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. J. Neurosci. 29, 2283–2296.
Lieberman, P.H., Klatt, D.H., Wilson, W.H., 1969. Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. Science 164, 1185–1187.
Linden, D.E., Thornton, K., Kuswanto, C.N., Johnston, S.J.V., v.d, V., Jackson, M.C., 2011. The Brain's voices: comparing nonclinical auditory hallucinations and imagery. Cereb. Cortex 21, 330–337.
Marx, V., 2016. Neurobiology: learning from marmosets. Nat. Methods 13, 911–916.
Meyer, M., Zysset, S., von Cramon, D.Y., Alter, K., 2005. Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. Cognitive Brain Res. 24, 291–306.
Meyers, E.M., Borzello, M., Freiwald, W.A., Tsao, D., 2015. Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. J. Neurosci. 35, 7069–7081.
Miller, C.T., Wren Thomas, A., 2012. Individual recognition during bouts of antiphonal calling in common marmosets. J. Comp. physiology. A, Neuroethol. Sens. neural, Behav. physiology 198, 337–346.
Miller, C.T., Mandel, K., Wang, X., 2010a. The communicative content of the common marmoset phee call during antiphonal calling. Am. J. Primatol. 72, 974–980.
Miller, C.T., Dimauro, A., Pistorio, A., Hendry, S., Wang, X., 2010b. Vocalization induced CFos expression in marmoset cortex. Front. Integr. Neurosci. 4, 128.
Miller, C.T., Freiwald, W.A., Leopold, D.A., Mitchell, J.F., Silva, A.C., Wang, X., 2016. Marmosets: a neuroscientific model of human social behavior. Neuron 90, 219–233.
Mullennix, J.W., Johnson, K.A., Topcu-Durgun, M., Farnsworth, L.M., 1995. The perceptual representation of voice gender. J. Acoust. Soc. Am. 98, 3080–3095.
Nagarajan, S.S., Cheung, S.W., Bedenbaugh, P., Beitel, R.E., Schreiner, C.E., Merzenich, M.M., 2002. Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J. Neurophysiol. 87, 1723–1737.
Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., Kojima, S., 2001. Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia 39, 1047–1054.
Newman, J.D., Kenkel, W.M., Aronoff, E.C., Bock, N.A., Zametkin, M.R., Silva, A.C., 2009. A combined histological and MRI brain atlas of the common marmoset monkey, Callithrix jacchus. Brain Res. Rev. 62, 1–18.
Nishimura, M., Takemoto, M., Song, W.J., 2018. Organization of auditory areas in the superior temporal gyrus of marmoset monkeys revealed by real-time optical imaging. Brain Struct. Funct. 223, 1599–1614.
Nummela, S.U., Jovanovic, V., de la Mothe, L., Miller, C.T., 2017. Social context-dependent activity in marmoset frontal cortex populations during natural conversations. J. Neurosci.
Okano, H., Sasaki, E., Yamamori, T., Iriki, A., Shimogori, T., Yamaguchi, Y., Kasai, K., Miyawaki, A., 2016. Brain/minds: a Japanese national brain project for marmoset neuroscience. Neuron 92, 582–590.
Ortiz-Rios, M., Kusmierek, P., DeWitt, I., Archakov, D., Azevedo, F.A., Sams, M., Jaaskelainen, I.P., Keliris, G.A., Rauschecker, J.P., 2015. Functional MRI of the vocalization-processing network in the macaque brain. Front. Neurosci. 9, 113.
Owren, M.J., Dieter, J.A., Seyfarth, R.M., Cheney, D.L., 1993. Vocalizations of rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques cross-fostered between species show evidence of only limited modification. Dev. Psychobiol. 26,

389–406.
Papoti, D., Yen, C.C., Mackel, J.B., Merkle, H., Silva, A.C., 2013. An embedded four-channel receive-only RF coil array for fMRI experiments of the somatosensory pathway in conscious awake marmosets. NMR Biomed. 26, 1395–1402.
Papoti, D., Yen, C.C., Hung, C.C., Ciuchta, J., Leopold, D.A., Silva, A.C., 2017. Design and implementation of embedded 8-channel receive-only arrays for whole-brain MRI and fMRI of conscious awake marmosets. Magn. Reson Med. 78, 387–398.
Pernet, C.R., Belin, P., 2012. The role of pitch and timbre in voice gender categorization. Front. Psychol. 3, 23.
Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. NEUROIMAGE 119, 164–174.
Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I., 2011. Voice cells in the primate temporal lobe. Curr. Biol. https://doi.org/10.1016/j.cub.2011.07.028.
Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I., 2014. Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. J. Neurosci. 34, 2524–2537.
Perrodin, C., Kayser, C., Abel, T.J., Logothetis, N.K., Petkov, C.I., 2015. Who is That? Brain networks and mechanisms for identifying individuals. Trends Cogn. Sci. 19, 783–796.
Petersen, M.R., Beecher, M.D., Zoloth, S.R., Moody, D.B., Stebbins, W.C., 1978. Neural lateralization of species-specific vocalizations by Japanese macaques (Macaca fuscata). Science 202, 324–327.
Petersen, M.R., Beecher, M.D., Zoloth, S.R., Green, S., Marler, P.R., Moody, D.B., Stebbins, W.C., 1984. Neural lateralization of vocalizations by Japanese macaques: communicative significance is more important than acoustic structure. Behav. Neurosci. 98, 779–790.
Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. Nat. Neurosci. 11, 367–374.
Petkov, C.I., Kikuchi, Y., Milne, A.E., Mishkin, M., Rauschecker, J.P., Logothetis, N.K., 2015. Different forms of effective connectivity in primate frontotemporal pathways. Nat. Commun. 6, 6000.
Pistorio, A.L., Vintch, B., Wang, X., 2006. Acoustic analysis of vocal development in a New World primate, the common marmoset (Callithrix jacchus). J. Acoust. Soc. Am. 120, 1655–1670.
Poremba, A., Malloy, M., Saunders, R.C., Carson, R.E., Herscovitch, P., Mishkin, M., 2004. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427, 448–451.
Rauschecker, J.P., 1998. Cortical processing of complex sounds. Curr. Opin. Neurobiol. 8, 516–521.
Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724.
Recanzone, G.H., 2008. Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. J. Neurosci. 29, 13184–13193.
Rendall, D., Rodman, P.S., Edmond, R.E., 1996. Vocal recognition of individuals and kin in free-ranging rhesus monkeys. Anim. Behav. 51, 1007–1015.
Rilling, J.K., 2014a. Comparative primate neuroimaging: insights into human brain evolution. Trends Cogn. Sci. 18, 46–55.
Rilling, J.K., 2014b. Comparative primate neurobiology and the evolution of brain language systems. Curr. Opin. Neurobiol. 28, 10–14.
Rinne, T., Muers, R.S., Salo, E., Slater, H., Petkov, C.I., 2017. Functional imaging of audio-visual selective attention in monkeys and humans: how do lapses in monkey performance affect cross-species correspondences? Cereb. Cortex 27, 3471–3484.
Romanski, L.M., Averbeck, B.B., 2009. The primate cortical auditory system and neural representation of conspecific vocalizations. Annu. Rev. Neurosci. 32, 315–346.
Romanski, L.M., Averbeck, B.B., Diltz, M., 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J. Neurophysiol. 93, 734–747.
Roswandowitz, C., Mathias, S.R., Hintz, F., Kreitewolf, J., Schelinski, S., von Kriegstein, K., 2014. Two cases of selective developmental voice-recognition impairments. Curr. Biol. 24, 2348–2353.
Rowell, T.E., Hinde, R.A., 1962. Vocal communication by the rhesus monkey (Macaca Mulatta). J. Zoology 138, 279–294.
Roy, S., Zhao, L., Wang, X., 2016. Distinct neural activities in premotor cortex during natural vocal behaviors in a new world primate, the common marmoset (Callithrix jacchus). J. Neurosci. 36, 12168–12179.
Russell, R., Duchaine, B., Nakayama, K., 2009. Super-recognizers: people with extraordinary face recognition ability. Psychon. Bull. Rev. 16, 252–257.
Sadagopan, S., Temiz-Karayol, N.Z., Voss, H.U., 2015. High-field functional magnetic resonance imaging of vocalization processing in marmosets. Sci. Rep. 5, 10950.
Schweinberger, S.R., Kawahara, H., Simpson, A.P., Skuk, V.G., Zaske, R., 2014. Speaker perception. Wiley interdisciplinary reviews. Cognitive Sci. 5, 15–25.
Silva, A.C., 2017. Anatomical and functional neuroimaging in awake, behaving marmosets. Dev. Neurobiol. 77, 373–389.
Snowdon, C.T., 2017. Learning from monkey "talk". Science 355, 1120–1122.
Suied, C., Agus, T.R., Thorpe, S.J., Mesgarani, N., Pressnitzer, D., 2014. Auditory gist: recognition of very short sounds from timbre cues. J. Acoust. Soc. Am. 135, 1380–1391.
Takahashi, D.Y., Fenley, A.R., Teramoto, Y., Narayanan, D.Z., Borjon, J.I., Holmes, P., Ghazanfar, A.A., 2015. LANGUAGE DEVELOPMENT. The developmental dynamics of marmoset monkey vocal production. Science 349, 734–738.
Talkington, W.J., Rapuano, K.M., Hitt, L.A., Frum, C.A., Lewis, J.W., 2012. Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. J. Neurosci. official J. Soc. Neurosci. 32, 8084–8093.
Teufel, C., Ghazanfar, A.A., Fischer, J., 2010. On the relationship between lateralized brain function and orienting asymmetries. Behav. Neurosci. 124, 437–445.
Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293.
Titze, I.R., 1989. Physiologic and acoustic differences between male and female voices. J. Acoust. Soc. Am. 85, 1699–1707.
Toarmino, C.R., Wong, L., Miller, C.T., 2017. Audience affects decision-making in a marmoset communication network. Biol. Lett. 13.
Toarmino, C.R., Yen, C.C., Papoti, D., Bock, N.A., Leopold, D.A., Miller, C.T., Silva, A.C., 2017. Functional magnetic resonance imaging of auditory cortical fields in awake marmosets. NEUROIMAGE 162, 86–92.
Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., Livingstone, M.S., 2006. A cortical region consisting entirely of face-selective cells. Science 311, 670–674.
Turesson, H.K., Ribeiro, S., Pereira, D.R., Papa, J.P., de Albuquerque, V.H., 2016. Machine learning algorithms for automatic classification of marmoset vocalizations. PLoS One 11, e0163041.
Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q. J. Exp. Psychol. A 43, 161–204.
Van Lancker, D., Kreiman, J., 1987. Voice discrimination and recognition are separate abilities. Neuropsychologia 25, 829–834.
Van Lancker, D.R., Cummings, J.L., Kreiman, J., Dobkin, B.H., 1988. Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex 24, 195–209.
Vanduffel, W., Zhu, Q., Orban, G.A., 2014. Monkey cortex through fMRI glasses. Neuron 83, 533–550.
Von Kriegstein, K., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NEUROIMAGE 22, 948–955.
Vouloumanos, A., Hauser, M.D., Werker, J.F., Martin, A., 2010. The tuning of human neonates' preference for speech. Child. Dev. 81, 517–527.
Wang, X., 2000. On cortical coding of vocal communication sounds in primates. Proc. Natl. Acad. Sci. U. S. A. 97, 11843–11849.
Wang, X., Kadia, S.C., 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J. Neurophysiol. 86, 2616–2620.
Wang, X., Merzenich, M.M., Beitel, R., Schreiner, C.E., 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J. Neurophysiology 74, 2685–2706.
Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., Petkov, C.I., 2013. Auditory artificial grammar learning in macaque and marmoset monkeys. J. Neurosci. 33, 18825–18835.
Winter, P., Funkenstein, H.H., 1973. The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (Saimiri sciureus). Exp. Brain Res. 18, 489–504.
Wollberg, Z., Newman, J.D., 1972. Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. Science 175, 212–214.
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 Lang. 149, 106–117.
Yovel, G., Belin, P., 2013. A unified coding strategy for processing faces and voices. Trends Cogn. Sci. 17, 263–271.
Zäske, R., Awwad Shiekh Hasan, B., Belin, P., 2017. It doesn't matter what you say: fMRI correlates of voice learning and recognition independent of speech content. Cortex 94, 100–112.
Zoloth, S.R., Petersen, M.R., Beecher, M.D., Green, S., Marler, P., Moody, D.B., Stebbins, W., 1979. Species-specific perceptual processing of vocal sounds by monkeys. Science 204, 870–873.