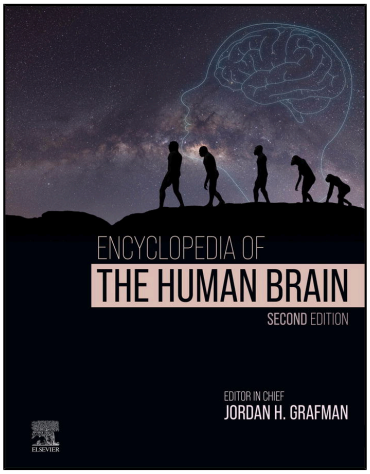
The cerebral architecture of voice information processing

Pascal Belin



Belin, P., 2025. The cerebral architecture of voice information processing. In: Grafman, J.H. (Ed.), Encyclopedia of the Human Brain, Second Edition, vol. 1, pp. 642–648.
USA: Elsevier.
ISBN: 9780128204801

The cerebral architecture of voice information processing

Pascal Belin, Institut de Neurosciences de la Timone, CNRS & Aix-Marseille Université, Marseille, France

Key points

- Voices are information-rich "auditory faces" that the human brain is expert at processing
- The three main types of voice information—speech, identity, affect—are processed in interacting, but dissociable functional pathways (as for faces)
- The Temporal Voice Areas (TVAs) constitute crucial nodes in the network of cortical regions processing voice information
- The TVAs categorize conspecific vocalizations apart from other sounds
- Evidence of TVAs in the rhesus macaque suggests a long evolutionary history of the vocal brain

Abstract

The human voice is arguably the most important sound category of our auditory environment by the wealth of socially-relevant information it carries: speech, but also identity and affect. The cerebral processing of voice information involves bilateral "temporal voice areas" (TVAs) that categorize voice apart from other sounds, and represent voice stimuli as a function of their acoustical dissimilarity with an average voice template. The finding of functionally homologous TVAs in the macaque brain suggests a long evolutionary history of the vocal brain.

Introduction

Imagine you are sitting on a train, and you hear a conversation in a foreign language in the row behind you. You do not see the speakers' bodies or faces, and you cannot understand the speech content because you do not know the language. Still, an amazing amount of information is available to you. You can evaluate the physical characteristics of the different protagonists, including their gender, approximate age and size, and associate an identity to the different voices. You can form a good idea of the different speaker's mood and affective state, and you also attribute personality traits to the speakers such as trustworthiness or dominance. In brief, despite the fact that linguistic information is unavailable, you can form a fairly detailed picture of the social interaction unfolding, which a brief glance backwards can on the occasion help refine—sometimes surprisingly.

This example is meant to illustrate our exquisite ability to extract and process the multiple types of information carried by voice—not "just" speech. It indicates the existence in the human brain of neural mechanisms dedicated to extracting and processing voice information. Unfortunately, most of the research effort directed at understanding these cerebral mechanisms has focused on a single type of voice information—speech. This is understandable given the uniqueness of human speech in the animal kingdom, but has resulted in a comparative lack of understanding of how other types of voice information are processed. The present article aims at presenting a non-exhaustive summary of research dedicated in the past two decades to understanding how the different types of voice information are processed by our brain.

The "auditory face" model of cerebral voice processing

Voices most often carry linguistic information ("speech"), although there are many instances in which voices don't carry any speech at all, as in laughs or coughs. But in addition (or not) to speech, voices also carry other types of information that also play a crucial role in our social interactions: person-related information on the speaker who produced the vocalization. We can perceive in voice

Encyclopedia of the Human Brain, Second Edition, 2025, 642–648

information on the stable physical characteristics of the speaker: his or her sex, her approximate age, an estimate of her size; we can form "vocal signatures" of speakers encountered repeatedly, allowing us to recognize them from novel utterances. We can also perceive in voice more transient aspects of the speaker's physiological state, allowing inferences on his or her motivational and emotional state. We also form more subtle impressions of a speaker's personality form his or her voice, that sometimes leave us confused when we see the actual face of the person. Thus, speech is only one of the multiple types of information a voice carry.

Remarkably, the different types of information we perceive in voice are also routinely perceived in faces. Despite the very different nature of their physical structure (light reflections hitting the retina in the eye vs. pressure waves inducing vibrations of the basilar membrane in the ear), faces and voices carry similar types of socially relevant information: both carry linguistic information (phonemes for voice, visemes for faces, i.e., representational units used to classify speech in the visual domain) but also relevant information on the identity and effective state of the speaker. From that perspective, voices can be considered as "auditory faces", and it is tempting to hypothesize that the analogy could extend to the underlying neuronal mechanisms. Indeed, the computational problems faced by the brain (robustness to noise, template matching, invariance, …) are of a similar nature for faces and voices, and the computational units (cortical columns) the same. This in turn suggests that the rich theoretical and methodological body of work on face perception can be used to generate hypotheses on the cognitive and cerebral mechanisms of voice perception.

Accordingly, we conceptualized the above notion by extending Bruce and Young's (1986) seminal model of face processing (Bruce and Young, 1986; see also Young and Bruce, 2011) and proposed a broadly similar functional architecture for voice processing (Belin et al., 2004), as already suggested by several authors (Ellis, 1989) (**Fig. 1**)—see Young et al. (2020) for a recent reformulation emphasizing differences in processing. This model is wrong, like all models, but the hope is that it can be useful in generating testable hypotheses.

The most basic, but fundamental information conveyed by voices is that they consist of human voices, i.e., sounds produced by the vocal tract of a conspecific human being. The "auditory face" model proposes a first processing stage where voices are perceptually categorized apart from all other sounds and represented relative to voice-specific templates: the "structural encoding" stage. After that initial voice-specific stage, the three main types of vocal information—speech, identity, affect—are further processed in three interacting, but partially dissociable functional pathways: (1) a pathway for analysis of linguistic (speech) information, involving anterior and posterior superior temporal sulcus (STS) as well as inferior prefrontal regions and pre-motor cortex predominantly in the left hemisphere; (2) a pathway for analysis of vocal affective information, involving temporo-medial regions, anterior insula, and amygdala and inferior prefrontal regions predominantly in the right hemisphere; (3) a pathway for analysis of vocal identity, involving "voice recognition units"—probably instantiated in regions of the right anterior STS—each activated by one of the voices known to the person (**Fig. 1**).

These three functional pathways are proposed to interact with each other during normal processing. For instance, interactions between the linguistic and identity pathways are well established: speech in noise is more intelligible if spoken by a familiar voice (Nygaard and Pisoni, 1998; Holmes et al., 2018), and conversely, speaker recognition is more robust when a familiar language is spoken—the "language familiarity effect" (Perrachione and Wong, 2007).

The model further proposes that a functional pathway can be selectively impaired while the two others function normally. That prediction is verified in the case of receptive aphasia, in which a patient experiences selective deficits in voice linguistic analysis while the perception of vocal identity or emotion remains normal; and in the case of phonagnosia, either acquired (Van Lancker et al.,

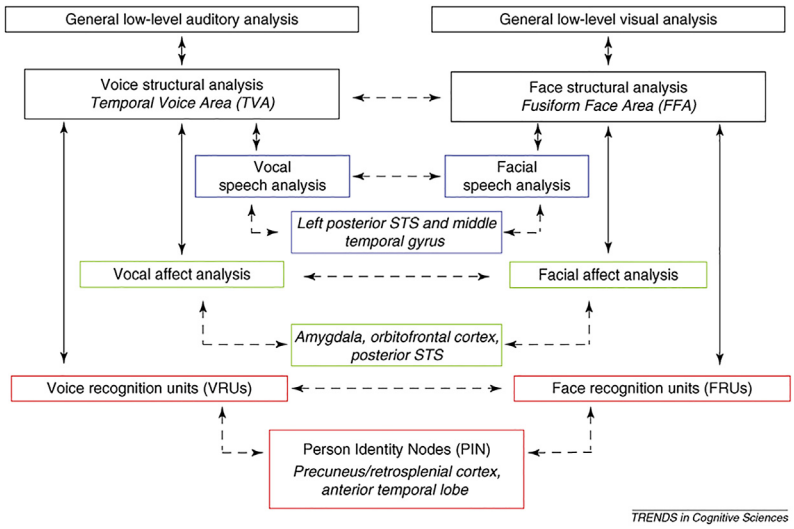


Fig. 1 **The "auditory face" model of voice perception.** The right-hand part of the figure is adapted from Bruce and Young's model of face perception (Bruce and Young, 1986). The left-hand part proposes a similar functional organization for voice processing. Dashed arrows indicate multimodal interactions. Candidate brain areas are proposed for each processing stream. Reproduced with permission from Campanella and Belin (2007).

1988) or developmental (Garrido et al., 2009), in which a person with normal speech comprehension and perception of voice affect is selectively impaired in recognizing speaker identity. Other predictions of the model (e.g., interaction between identity and affective pathways, or existence of patients with selective deficits in voice affect perception) remain to be challenged.

The temporal voice areas (TVAs) and the extended voice network

Functional MRI studies comparing cortical responses to vocal vs. nonvocal sounds show that the human cerebral substrate for voice perception is centered on secondary auditory cortical regions located bilaterally along the superior temporal gyrus (STG) and sulcus (STS). These "temporal voice areas" (TVAs) show greater fMRI signal in response to vocal sounds, whether they contain speech or not, compared to other categories of non-vocal sounds such as environmental sounds, amplitude-modulated noise (Belin et al., 2000, 2002; Von Kriegstein and Giraud, 2004; Grandjean et al., 2005; Ethofer et al., 2007; Ethofer et al., 2009; Blank et al., 2011; Linden et al., 2011; Moerel et al., 2013), or to hetero-specific vocalisations (HV) (Fecteau et al., 2004). Sound stimuli including vocal and nonvocal sounds necessary for a "voice localizer" fMRI scan can be found here: https://neuralbasesofcommunication.eu/download/.

The TVAs are fairly variable in exact anatomical location across individuals and hemispheres. Yet they are highly reliable within subject as indicated by test-retest reliability analysis (Pernet et al., 2015). A mega-analysis of the voice localizer collected in several hundred participants (Pernet et al., 2015) included a cluster analysis of local voice-sensitivity maxima that suggested an organization in three "voice patches" along the antero-posterior axis of the STG and STS bilaterally. And indeed, 3 bilateral maxima of voice selectivity can be found in most individuals along STG/STS: we call these 3 voice patches the posterior, middle and anterior TVAs (pTVA, mTVA and aTVA). One important unanswered question concerns the exact functional differences between these three voice patches: as for faces, they could instantiate increasingly abstract, invariant representations of the vocal signal.

The power afforded by the large sample size in Pernet et al. (2015) also showed that the TVAs are but the most salient part of an extended "vocal brain", a bilateral, distributed network of cortical and subcortical regions showing small but significant voice-sensitivity. As for face processing (Haxby et al., 2000), evidence suggests the TVAs constitute a Core network for voice processing in the temporal lobe connected to an Extended network comprised of additional cortical and subcortical areas. The Extended voice network notably includes 3 sets of bilateral prefrontal areas—the posterior, middle and anterior prefrontal Voice Areas (a, m and PFVA) (Aglieri et al., 2018) as well as the amygdala (**Fig. 2**).

Our voice cognition abilities most often involve more than simply detecting a voice, and our behavioral goals often make us focus on one specific type of voice information: speech content, identity, affective state, etc. Results from neuroimaging studies on the perception of voice gender (Charest et al., 2013), identity (Latinus et al., 2011, 2013), affect (Bestelmeyer et al., 2014b) or attractiveness (Bestelmeyer et al., 2012) converge to the notion of an interplay between the Core and Extended voice networks to extract and process voice information relevant to behavioral goals. These studies suggest that neuronal activity in the Core Network is essentially related to extracting voice acoustics, essentially driven by acoustical proximity to the voice template in the middle voice patches TVAm, and by acoustical differences with the previously heard voice stimulus in more anterior TVAa. That acoustical information is further processed by the Extended Network as a function of the task demands, involving subcortical areas such as the amygdala (Bestelmeyer et al., 2014a,b) and several areas of prefrontal cortex bilaterally (Bestelmeyer et al., 2012, 2014b; Charest et al., 2013; Latinus et al., 2011, 2013).

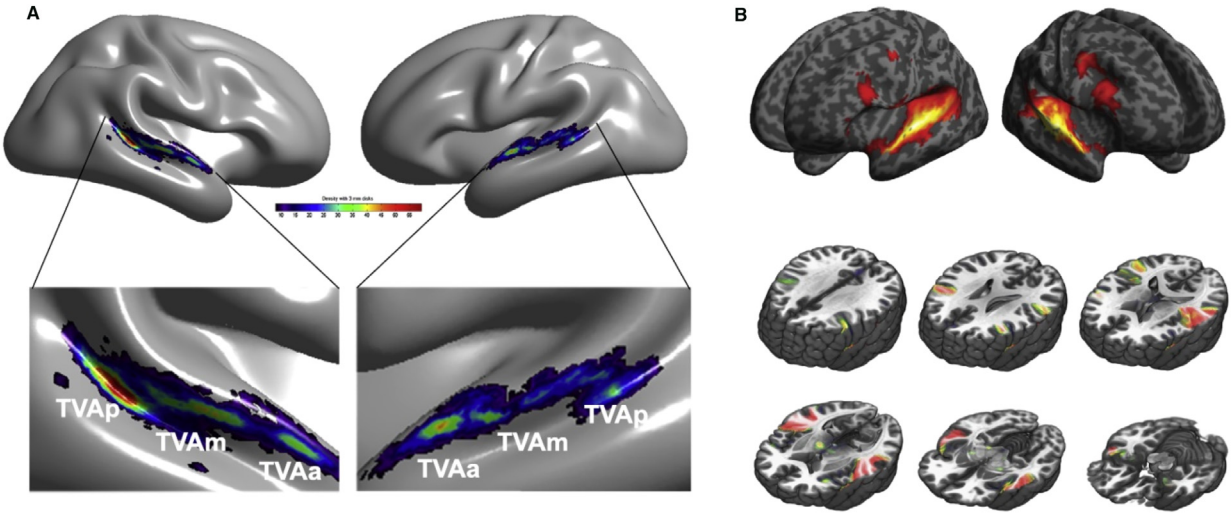


Fig. 2 **Core and extended voice networks.** (A) The Core Network: Three bilateral temporal voice areas (TVAs) as suggested by cluster analysis of individual voice sensitivity maxima. (B) The Extended Network: group analyses with large sample sizes reveals voice sensitivity in a number of extra temporal areas notably in bilateral inferior prefrontal cortex and the amygdala. Adapted from Pernet et al. (2015).

The existence of TVAs in the cortex of the vast majority of normal listeners has been verified by several groups and can hardly be disputed. What remains less clear is the functional role of the different voice patches. Are these cortical areas even performing computations related to voice perception? The TVA patches do not only respond to voices: they also respond to other complex sounds, although less strongly. One possibility is that these areas are part of a larger set of areas performing general timbre analysis—analogous to cortical areas performing general visual shape analysis of which the FFA and OFA constitute prominent nodes (Haxby et al., 2001). This potential analogy is reinforced by neuroimaging studies of timbre processing that show that areas of anterior temporal lobe close to the location of the central TVAm cluster respond to the manipulation of different timbre cues in complex sounds (Warren et al., 2005; Menon et al., 2002). Another clue on the TVA's functional role comes from studies of social cognition that indicate a striking involvement of neuronal populations along the STS for processing social signals in general—stimuli from other humans such as mouth, hand, eyes, etc. The observation of voice-sensitive neuronal populations along the STS, particularly in the most posterior voice patch TVAp, constitute additional evidence for a strong involvement of STS cortex in representing social stimuli.

One influential hypothesis on the functional significance of the FFA activations in response to faces if that they reflect a computational stage of face detection (Tsao and Livingstone, 2008). This face detection stage would constitute a mandatory computational stage before processing further facial information according to behavioral goals. Indeed automatic face processing research suggests it is computationally much more rapid and efficient to first detect face stimuli then apply specific fine-grained template matching or filter mechanisms only on the detected faces to extract higher-level face information (identity, affect, …) rather than to apply those filters or match internal templates directly on all incoming stimuli, faces as well as non-face objects (Tsao and Livingstone, 2008). Such mandatory face detection stage before more advanced goal-oriented processing is akin to what Vicky Bruce and Andy Young conceptualize as a face "structural encoding" stage in their influential model of face processing (Bruce and Young, 1986; Young and Bruce, 2011): a processing stage where a stimulus is recognized as a face and at which its internal structure is encoded for further processing along three main processing routes.

The analogy between face and voice processing suggests that the notion of a mandatory detection stage could also apply for the TVAs. The "voice structural encoding" stage of the "Auditory Face" voice processing model would correspond to voice detection, i.e., the recognition that an incoming sound stimulus is a human voice. As for faces this stage could consist of a mandatory processing stage before higher-level processing of different types of voice information.

This "structural encoding" stage could not only detect a voice but also at the same time match it to internal voice templates. This notion is consistent with evidence that the TVA encode novel voices using a norm-based code (Latinus et al., 2013): voices more acoustically different from an internal voice template (actually two: one male and one female) are perceived as more distinctive and elicit greater activity in the TVAs than voices more similar to the prototype. This encoding scheme, strikingly similar to face encoding schemes (Leopold et al., 2006; Loffler et al., 2005; Koyano et al., 2021) would constitute a parsimonious way of encoding voice information for further processing in the extended voice network.

Evolution of the TVAs

How the evolution of speech has transformed the human auditory cortex compared to other primates remains largely unknown. While primary auditory cortex is organized largely similarly in humans and macaques (Kaas and Hackett, 2000) the picture is much less clear at higher levels of the anterior auditory pathway (Rauschecker et al., 1995), particularly regarding the processing of conspecific vocalizations (CVs).

A "voice region" similar to the human TVAs has been identified in the macaque right anterior temporal lobe with functional MRI (Petkov et al., 2008); however its anatomical localization seemingly inconsistent with that of the human TVAs has suggested a "repositioning of the voice area" in recent human evolution (Ghazanfar, 2008). A more recent comparative fMRI study observed important differences in the cerebral processing of harmonic sounds in the anterior temporal lobe of humans and macaques (Norman-Haignere et al., 2019), suggesting a "fundamental divergence" in the organization of higher-level auditory cortex.

Both studies suggest that the cerebral processing of CVs (often highly harmonic) might have fundamentally diverged in the human lineage compared to other primates; they cast doubts on the usefulness of non-human primate models for studying higher-level audition in humans. However no study yet has compared vocalization processing by humans and macaques using the same MRI scanner and experimental protocol and with conclusive results at standard significance thresholds (but see Joly et al. (2012)).

To address this issue we recently used comparative fMRI and scanned awake rhesus macaques ($n = 3$) and humans ($n = 5$) on the same 3T MRI scanner (Siemens Prisma) using an identical auditory stimulation paradigm. Auditory stimuli ($n = 96$) consisted of brief complex sounds sampled from 16 categories grouped in 4 larger categories: human speech and voice ($n = 24$), macaque vocalizations ($n = 24$), marmoset vocalizations ($n = 24$) as well as complex non-vocal sounds ($n = 24$). The comparison of fMRI volumes acquired during sound stimulation vs. the silent baseline revealed general auditory activation by the stimulus set. Both humans and macaques showed extensive bilateral STG activation centered in both species on core areas of the auditory cortex (A1) and extending rostrally and caudally to higher-level auditory cortex.

We next searched for CV-selective activations by contrasting in each species the fMRI signal measured in response to CVs vs. all other sounds. In humans this comparison confirmed the classical pattern of three TVAs along mid-STS to anterior STG bilaterally with additional voice-selective activations in premotor and inferior frontal cortex including in bilateral BA 44 and 45. In macaques,

the contrast of macaque vocalizations vs. all other sounds primarily yielded bilateral CV-selective activations in anterior STG and extending ventrally to the upper bank of STS in the left hemisphere. These regions correspond to the cytoarchitectonic area ts2, confirming the localization of the macaque voice area previously described in the right hemisphere (Petkov et al., 2008) while emphasizing its bilateral nature: the left voice area was actually more consistently located in our macaques than its right counterpart. We call these bilateral voice areas the macaque aTVAs because of their anterior STG localization analogous to that of the human aTVAs (Bodin et al., 2021) (**Fig. 3**).

We then compared sound representations in the aTVA with those in primary auditory cortex (A1) using Representational Similarity Analysis (Kriegeskorte et al., 2009). In A1, we found strikingly similar sound representations in humans and macaques, essentially reflecting acoustical differences between stimuli (Bodin et al., 2021). In the aTVA, in contrast, sound representations were

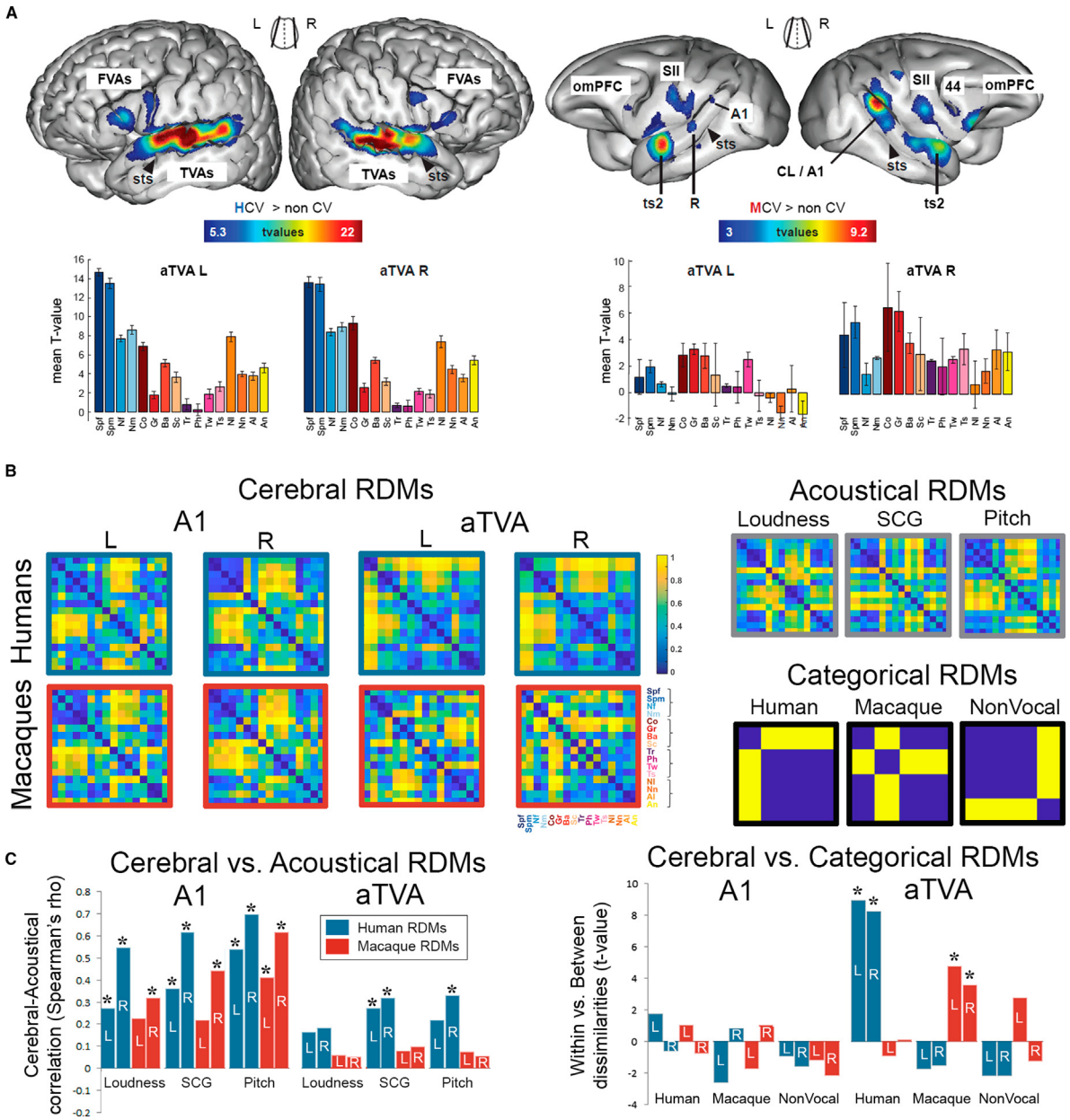


Fig. 3 **Functionally homologous aTVAs in humans and macaques.** (A) The contrast of fMRI signal acquired during stimulation with conspecific vocalizations vs. all other sounds highlights TVAs in humans (left) and macaques (right). Bar plots indicate the responses of the left and right aTVAs to the 16 subcategories of sound stimuli—4 leftmost categories in blue are human voices, 4 next in red are macaque vocalizations. (B) Left: Cerebral Representational Dissimilarity Matrices (RDMs) capturing pairwise differences in BOLD signal evoked by the 16 sound categories. Right: acoustical RDMs and categorical model RDMs for the 16 categories. (C) Comparison of the cerebral RDMs to the acoustical and model RDMs shows strong correlations with acoustics in A1 of both species, but strongest correlation with the species-specific categorical model in aTVAs. Adapted from Bodin et al. (2021).

markedly different in humans and macaques: in both species BOLD signal differences between sounds were best explained by a binary model categorizing conspecific vocalizations apart from the other sounds. Thus, not only have the aTVAs analogous anatomical localizations, they are functionally homologous in humans and macaques.

A group in Budapest recently developed a dog fMRI scanning strategy where dogs are trained using operant conditioning with positive food and social reinforcement to stay still in the scanner for about 10 min—which they learn much faster than the macaques! A group of dogs was scanned during auditory stimulation with sounds of dog vocalizations, human voices and environmental sounds, allowing the delineation of functionally sensitive auditory cortex and the comparison of selectivity in different cortical areas (Andics et al., 2014). The contrast of images of brain activity measured during stimulation with dog vocalizations vs. the other sounds highlights a small region of middle ectosylvian gyrus that shows significant preference of dogs vocalizations—a dog homolog of the voice areas. Furthermore that area was found to be modulated by the affective content of the vocalizations (Andics et al., 2014), much like the human TVAs (Ethofer et al., 2009, 2012).

Together the macaque and dog findings of TVA homologs provide strong evidence for a long evolutionary history of voice processing. The most parsimonious explanation for the fact that voice areas can be found today in these species is that the voice areas were already present in some rudimentary form in the last common ancestor of humans, macaques and dogs, more than 80 million years ago! Thus the neural mechanisms dedicating to analyzing conspecific vocalizations have been evolving during these tens of millions of years and it is not surprising that they are highly developed in our brain. This notion implies that we should be able to find voice areas in the brain of many other living animals for whom accurate processing of conspecific vocalizations has some adaptive significance. This also implies that than when they started speaking some 0.1−0.2 million years ago, our ancestors were already equipped with sophisticated neural machinery for processing voice information which they could bootstrap to develop a novel mode of vocal communication—speech.

Conclusion

Research into the cerebral mechanisms of the non-linguistic aspects of voice cognition has been lagging compared to speech or face perception research, but it is gaining increasing momentum. Current results suggest, in strong analogy with known mechanisms of cerebral face processing, a distributed vocal brain with a core network of areas centered on the TVAs and an extended network including extra-temporal areas such as the amygdala and areas of inferior prefrontal cortex. The core network corresponds to a stage of "voice structural encoding" that performs voice detection and acoustical template matching while the extended network performs higher-level goal-related computations organized in three interacting processing streams specialized in extracting one particular type of vocal information.

Many open questions remain for a better understanding of the functional architecture of the vocal brain. About the core network in the TVAs: How different is the vocal stimulus representation in the different voice patches? Can homologs of the voice patches be found in macaques and do they show a progression in the voice representation similar to what has been evidenced for the macaque face patches (Freiwald and Tsao, 2010; Freiwald et al., 2009)? Are there precise anatomo-functional correspondences in the human TVAs that could help interpret their large inter-individual variability, and could such correspondences be observed in simplified form in the macaque brain? About the extend network: what is the precise functional topography of prefrontal areas engaged by voices? What is their pattern of anatomical connectivity with the core TVAs and how does this co-vary with inter-individual differences in behavioral performance? How is the pattern of functional connectivity between core and extended networks modulated by task demands and how is it affected in acquired or developmental phonagnosia? etc.

Also, while voice perception and production are intimately related, the extent to which their neural substrates overlap or interact remains largely unexplored. Is activity in TVA modulated during vocalization production (in a similar fashion as some neurons in primary auditory cortex are inhibited)? What is the connectivity between the voice core and extended network and voice production areas? Are there differences in activity depending on whether subjects are passively listening to voices or if they are engaged in conversations? Much exciting work lies ahead on our road to answer these important questions on the functioning of one of our most ancient and important ability to extract information from conspecific vocalizations.

Acknowledgments

I thank the many students and colleagues who contributed to this work, and an anonymous reviewer for constructive comments. The work was supported by grant BB/E003958/1 from BBSRC (UK), large grant RES-060-25-0010 by ESRC/MRC (UK), grant AJE201214 by the Fondation pour la Recherche Médicale, Agence Nationale de la Recherche grants ANR-16-CE37-0011-01 (PRIMAVOICE), ANR-16-CONV-0002 (Institute for Language, Communication and the Brain), ANR-11-LABX-0036 (Brain and Language Research Institute), the Excellence Initiative of Aix-Marseille University (A*MIDEX), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 788240).

References

Aglieri, V., Chaminade, T., Takerkart, S., Belin, P., 2018. Functional connectivity within the voice perception network and its behavioural relevance. Neuroimage 183, 356–365.
Andics, A., Gacsi, M., Farago, T., Kis, A., Miklosi, A., 2014. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. 24, 574–578.
Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cognit. Brain Res. 13, 17–26.
Bestelmeyer, P.E., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., Belin, P., 2012. Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. Cerebr. Cortex 22, 1263–1270.
Bestelmeyer, P.E., Belin, P., Ladd, D.R., 2014a. A neural marker for social bias toward in-group accents. Cerebr. Cortex 25 (10), 3953–3961.
Bestelmeyer, P.E., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014b. Adaptation to vocal expressions reveals multistep perception of auditory emotion. J. Neurosci. 34, 8098–8105.
Blank, H., Anwander, A., von Kriegstein, K., 2011. Direct structural connections between voice- and face-recognition areas. J. Neurosci. 31 (36), 12906–12915.
Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., Rapha, E., Renaud, L., Giordano, B.L., Belin, P., 2021. Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques. Curr. Biol. 31, 4839–4844 e4.
Bruce, V., Young, A., 1986. Understanding face recognition. Br. J. Psychol. 77, 305–327.
Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. Trends Cognit. Sci. 11, 535–543.
Charest, I., Pernet, C., Latinus, M., Crabbe, F., Belin, P., 2013. Cerebral processing of voice gender studied using a continuous carryover FMRI design. Cerebr. Cortex 23, 958–966.
Ellis, A.W., 1989. Neuro-cognitive processing of faces and voices. In: Young, A.W., Ellis, H.D. (Eds.), Handbook of Research on Face Processing. Elsevier Science Publishers B.V., pp. 207–215
Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. Curr. Biol. 19, 1028–1033.
Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cerebr. Cortex 22, 191–200.
Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., Wildgruber, D., 2007. The voices of seduction: cross-gender effects in processing of erotic prosody. Soc. Cogn. Affect Neurosci. 2, 334–337.
Fecteau, S., Armony, J., Joanette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage 23, 840–848.
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.
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.
Ghazanfar, A.A., 2008. Language evolution: neural differences that make a difference. Nat. Neurosci. 11, 382–384.
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.
Haxby, J.V., Hoffman, E.A., Ida Gobbini, M., 2000. The distributed human neural system for face perception. Trends Cognit. Sci. 4, 223–233.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.
Holmes, E., Domingo, Y., Johnsrude, I.S., 2018. Familiar voices are more intelligible, even if they are not recognized as familiar. Psychol. Sci. 29, 1575–1583.
Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012. Processing of vocalizations in humans and monkeys: a comparative fMRI study. Neuroimage 62, 1376–1389.
Kaas, J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. U. S. A. 97, 11793–11799.
Koyano, K.W., Jones, A.P., Mcmahon, D.B.T., Waidmann, E.N., Russ, B.E., Leopold, D.A., 2021. Dynamic suppression of average facial structure shapes neural tuning in three macaque face patches. Curr. Biol. 31, 1–12 e5.
Kriegeskorte, N., Mur, M., Bandettini, P., 2009. Representational similarity analysis—connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (4), 1–28.
Latinus, M., Crabbe, F., Belin, P., 2011. Learning-induced changes in the cerebral processing of voice identity. Cerebr. 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.
Leopold, D.A., Bondar, I.V., Giese, M.A., 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442, 572–575.
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 aauditory hallucinations and imagery. Cereb. Cortex 21, 330–337.
Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. fMRI evidence for the neural representation of faces. Nat. Neurosci. 8, 1386–1390.
Menon, V., Levitin, D.J., Smith, B.K., Lembke, A., Krasnow, B.D., Glazer, D., Glover, G.H., Mcadams, S., 2002. Neural correlates of timbre change in harmonic sounds. Neuroimage 17, 1742–1754.
Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., Formisano, E., 2013. Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. J. Neurosci. 33 (29), 11888–11898.
Norman-Haignere, S.V., Kanwisher, N., Mcdermott, J.H., Conway, B.R., 2019. Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. Nat. Neurosci. 22, 1057–1060.
Nygaard, L.C., Pisoni, D.B., 1998. Talker-specific learning in speech perception. Percept. Psychophys. 60, 355–376.
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.
Perrachione, T.K., Wong, P.C., 2007. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. Neuropsychologia 45, 1899–1910.
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.
Rauschecker, J.P., Tian, B., Hauser, M., 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111–114.
Tsao, D.Y., Livingstone, M.S., 2008. Mechanisms of face perception. Annu. Rev. Neurosci. 31, 411–437.
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.
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.
Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24, 1052–1057.
Young, A.W., Bruce, V., 2011. Understanding person perception. Br. J. Psychol. 102, 959–974.
Young, A.W., Fruhholz, S., Schweinberger, S.R., 2020. Face and voice perception: understanding commonalities and differences. Trends Cognit. Sci. 24, 398–410.