



HAL
open science

Similarities in face and voice cerebral processing

Pascal Belin

► **To cite this version:**

Pascal Belin. Similarities in face and voice cerebral processing. *Visual Cognition*, 2017, 25 (4-6), pp.658 - 665. 10.1080/13506285.2017.1339156 . hal-04687248

HAL Id: hal-04687248

<https://hal.science/hal-04687248v1>

Submitted on 4 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Co-Editors: James W. Tanaka & Christian N.L. Olivers

Special issue:
Person Perception
Guest editors:
Gail Yovel, Galia Avidan, and Alice O'Toole



ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: www.tandfonline.com/journals/pvis20

Similarities in face and voice cerebral processing

Pascal Belin

To cite this article: Pascal Belin (2017) Similarities in face and voice cerebral processing, Visual Cognition, 25:4-6, 658-665, DOI: [10.1080/13506285.2017.1339156](https://doi.org/10.1080/13506285.2017.1339156)

To link to this article: <https://doi.org/10.1080/13506285.2017.1339156>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 05 Jul 2017.



Submit your article to this journal [↗](#)



Article views: 2089



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

Similarities in face and voice cerebral processing

Pascal Belin  ^{a,b,c}

^aInstitut de Neurosciences de La Timone, UMR 7289 CNRS and Aix-Marseille University, Marseille, France; ^bDépartement de Psychologie, Université de Montréal, Montréal, Québec, Canada; ^cInstitute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, UK

ABSTRACT

In this short paper I illustrate by a few selected examples several compelling similarities in the functional organization of face and voice cerebral processing: (1) Presence of cortical areas selective to face or voice stimuli, also observed in non-human primates, and causally related to perception; (2) Coding of face or voice identity using a “norm-based” scheme; (3) Personality inferences from faces and voices in a same Trustworthiness–Dominance “social space”.

ARTICLE HISTORY

Received 31 October 2016
Accepted 3 May 2017

KEYWORDS

Face; voice; identity; selectivity; attractiveness

Although the nature of the sensory input is highly different for facial or vocal information, growing evidence suggests that the cerebral architecture processing these two types of signals is organized following similar principles (Yovel & Belin, 2013). This short paper provides a biased, non-exhaustive comparison of the perceptual and neural mechanisms involved in face and voice processing, focusing on a few examples, mostly from my own work, that illustrate their puzzling similarities.

1. Neural selectivity: faces and voices are special

It is now well established that faces are “special”: combined evidence from cognitive psychology experiments, studies in brain lesioned patients, and electrophysiology and neuroimaging studies in primates and humans (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bodamer, 1947; Bruce & Young, 1986; Haxby, Hoffman, & Ida Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; Sergent & Signoret, 1992; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Tsao, Freiwald, Tootell, & Livingstone, 2006; Young & Bruce, 2011) all point to the fact that viewing a face engages psychological and neural mechanisms not engaged by other object categories. Are voices “special” too? It would seem that they are.

- a. *Human neuroimaging.* Functional magnetic resonance imaging (fMRI) studies in humans have

evidenced Temporal Voice Areas (TVAs) in human auditory cortex (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) analogous to the “face areas” or “face patches” of visual cortex (Freiwald, Tsao, & Livingstone, 2009; Haxby, Hoffman, & Ida Gobbini, 2000; Kanwisher et al., 1997; Tsao et al., 2006). TVAs show greater response to voices – whether they contain speech or not – than to other categories of nonvocal sounds from the environment or to acoustical control stimuli such as scrambled voices and amplitude-modulated noise. They are organized bilaterally in several clusters along the superior temporal gyrus and superior temporal sulcus of the temporal lobe (Belin et al., 2000; Belin, Zatorre, & Ahad, 2002; Linden, Thornton, Kuswanto, Johnston, & Jackson, 2011; Von Kriegstein & Giraud, 2004). A recent large analysis of cerebral voice sensitivity in several hundred participants (Pernet et al., 2015) demonstrates that TVAs are the most salient part of a “vocal brain”, a bilateral, distributed network of cortical and subcortical regions showing significant voice-sensitivity including in particular inferior prefrontal areas and the amygdala. A cluster analysis of peaks of voice sensitive response in these hundreds of participants provided evidence of an organization in three “voice patches” along the antero-posterior axis of the superior temporal sulci and gyri bilaterally (Figure 1).

CONTACT Pascal Belin  pascal.belin@univ-amu.fr

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- b. *Causal link with perception.* The increased response to faces seen in face-sensitive cortical areas is causally related to face processing. Transiently interfering with neuronal populations in the occipital face area using transcranial magnetic stimulation (TMS) results in specific drops in behavioural performance tasks involving the processing of faces, but not other objects (Pitcher, Charles, Devlin, Walsh, & Duchaine, 2009; Pitcher, Walsh, Yovel, & Duchaine, 2007). Likewise, TMS stimulation of the TVA interferes with voice perception (Bestelmeyer, Belin, & Grosbras, 2011). Repetitive TMS stimulation of the TVA peak in the right hemisphere induces a performance level difference between a voice perception task (voice/nonvoice categorization) and a low-level nonvocal auditory task (loudness judgment) that is not observed when stimulating a control site in parietal cortex (Bestelmeyer et al., 2011). Although that study is but the starting point of a line of research potentially as productive as that using TMS to dissect face processing, it clearly establishes a causal link between TVA activation and voice perception, as found for OFA activation and face processing.
- c. *Human electrophysiology.* Electrophysiological techniques such as electro- (EEG) or magnetoencephalography (MEG) reveal comparable time courses for face and voice processing. For faces, both techniques reveal a well-known N170/N170 m component most prominent on occipito-temporal electrodes bilaterally with generally higher amplitude in response to faces than to other objects (Bentin et al., 1996). Similarly, high-density EEG shows that bilateral fronto-temporal electrodes display a component in the P200 range called the “fronto-temporal positivity to voices” (FTPV) with a larger amplitude in response to vocal compared to nonvocal sounds as early as about 170 msec after sound onset (Charest et al., 2009). MEG confirms this finding and identifies sources of the FTPVm in anterior/posterior STG/STS bilaterally, overlapping with the fMRI-derived anatomical location of TVAs (Capilla, Belin, & Gross, 2013). Thus, it takes about two-tenths of a second for our brain to differentiate face or voice from other signals in the same sensory modality.
- d. *Non-human studies.* fMRI studies face processing in macaques have consistently identified a series

of cortical “face patches” and, together with electrophysiology, described their functional properties to a more detailed degree than is feasible in humans; the results have suggested a series of increasingly detailed face representations as one moves anteriorly towards the frontal lobe, with some patches containing a large proportion of “face cells”, i.e., neurons displaying face preference at the individual level (Freiwald & Tsao, 2010; Freiwald et al., 2009; Tsao et al., 2006).

The current understanding of voice processing in the macaque brain is less advanced and different studies yield partly conflicting results with some groups not finding any reliable cortical areas differentiating conspecific vocalization from other complex sounds (Joly et al., 2012), while other groups show differences in variable cortical areas (Gil-da-Costa et al., 2004; Gil-da-Costa et al., 2006; Ortiz-Rios et al., 2015; Petkov et al., 2008). A pioneering study by Chris Petkov and colleagues provided clear evidence for TVAs in the macaque brain: using fMRI in awake macaques during auditory stimulation with macaque vocalizations and other sound categories they observed, in two animals, several “voice patches” in temporal lobe with greater activity in response to macaque vocalizations than to other complex sounds (Petkov et al., 2008) – the macaque equivalent of the human voice areas. One of the voice patches, located in right anterior temporal lobe, showed adaptation to speaker identity (Petkov et al., 2008), similar to evidence obtained in humans in an analogous anatomical location (Belin & Zatorre, 2003). Moreover, single cell recordings performed in this fMRI-identified location provided the first evidence of “voice cells”, i.e., individual neurons showing significant selectivity to conspecific vocalizations (Perrodin, Kayser, Logothetis, & Petkov, 2011). These results are important in that they strongly suggest that the last common ancestor of humans and macaques, some 20–25 million years ago, were already equipped with rudimentary voice-selective cortical mechanisms. Thus, when our hominin ancestors started speaking a few tens or hundred thousand years ago, they were already equipped with neural mechanisms tuned over millions of years to analyse voice information. Interestingly, fMRI studies in dogs have also recently provided evidence of voice areas in the dog brain: areas responding significantly more to dog

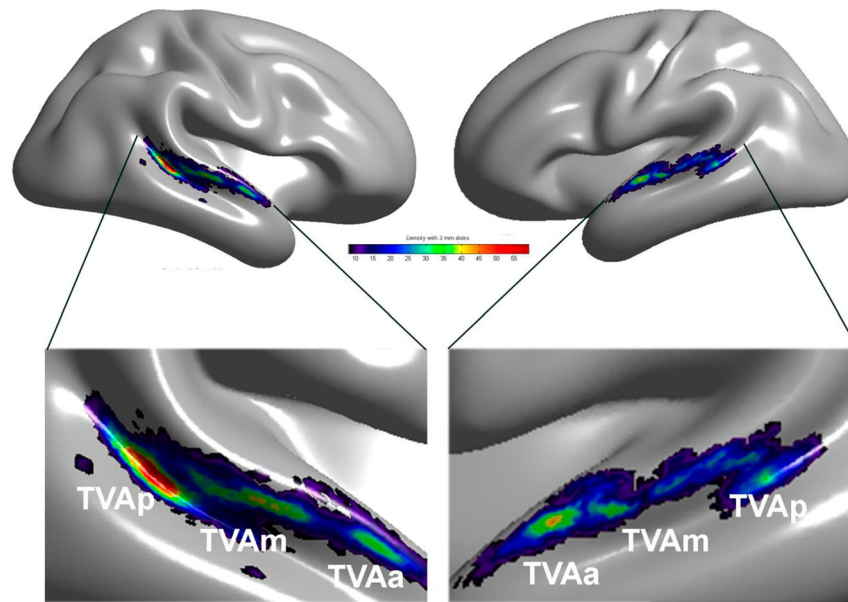


Figure 1. Temporal Voice Areas (TVAs) in the human brain. A cluster analysis of the density map of voice > nonvoice contrast revealing three main clusters of voice sensitivity in each hemisphere along a voice-sensitive zone of cortex extending from posterior STS to mid-STS/STG to anterior STG. The cluster with the greatest peak density is in right pSTS, consistent with individual images. Reproduced (permission pending) from Pernet et al. (2015).

vocalizations compared to other sounds (Andics, Gacsi, Farago, Kis, & Miklosi, 2014), which pushes back the emergence of the vocal brain to 80 million years ago. Future studies should confirm the existence of these voice patches, detail their anatomical location and inter-individual variability, and examine potential differences in underlying voice representations.

2. Identity processing

Faces and voices are the two most important signal categories allowing us to recognize other individuals. A large number of studies have investigated the functional and neuronal architecture underlying face recognition (Bruce & Young, 1986; Calder & Young, 2005; Tsao & Livingstone, 2008; Young & Bruce, 2011); comparatively less effort has been devoted to studying voice speaker recognition (Blank, Wieland, & von Kriegstein, 2014; Perrodin, Kayser, Abel, Logothetis, & Petkov, 2015), however what is known reveals troubling similarities with face recognition.

- a. *Selective recognition deficits.* Selective recognition deficits are known to occur for both face and voice identity processing. Since Bodamer (1947), many cases of “prosopagnosia” have been documented: patients who following a brain lesion become unable to recognize previously known

faces while still being able to recognize nonface object categories (Rossion, 2014; Sergent & Signoret, 1992). Some persons even present face recognition impairments in the absence of any evident brain damage – a deficit termed “developmental” or “congenital” prosopagnosia (Duchaine & Nakayama, 2006). Importantly other aspects of face perception, such as the ability to recognise emotions or lip-read, seem to be preserved in prosopagnosic patients, indicating that the functional pathway underlying face identity processing is partially dissociated from those underlying emotional or speech information processing (Bruce & Young, 1986; Young & Bruce, 2011).

A directly comparable deficit is also known to occur for speaker recognition – although documented in a much smaller number of cases (Van Lancker, Cummings, Kreiman, & Dobkin, 1988; Van Lancker, Kreiman, & Cummings, 1989). This deficit in speaker recognition, called “phonagnosia”, also occurs more frequently after right hemisphere lesions. As for prosopagnosia, the deficit seems quite selective to identity information processing as these patients typically show normal speech comprehension or emotion recognition from voice. A small number of cases of so-called “developmental phonagnosia”, presenting the

deficit quite selectively without evident brain damage, have also recently been described (Garrido et al., 2009; Roswadowitz et al., 2014; Xu et al., 2015). A detailed investigation of the neural correlates of such deficits is the subject of ongoing investigations by several groups.

- b. *Perceptual coding*. How is voice or face identity coded in the brain? In both cases there appears to be distinct mechanisms for the coding of identity in familiar or unfamiliar faces/voices. In the case of faces this is shown for instance by the good performance of typical observers to match different views of a familiar faces in tests such as the Benton test, whereas different views of a same unfamiliar face are often perceived as corresponding to different identities. For voices a comparable dissociation has been observed between discrimination of unfamiliar speakers versus recognition of known speakers (Van Lancker et al., 1988).

The coding of unfamiliar identities appears to be performed for both faces and voices using a Norm-based coding mechanism. For faces, at the behavioural level, adaptation after-effects cause larger identity categorization performance differences for face adapters that sit opposite the target face relative to the prototypical face so-called called “anti-faces” – which has been interpreted as highlighting the special role of the prototype in identity coding (Leopold, O’Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006). At the neuronal level, single cell recordings in macaque inferior temporal lobe (Leopold, Bondar, & Giese, 2006) as well as human fMRI measures of FFA activity (Loffler, Yourganov, Wilkinson, & Wilson, 2005) indicate that faces more dissimilar to an identity-free face prototype (approximated by computer averaging of many different faces) elicit greater neuronal activity than less distinctive faces, more similar to the prototype. For voices, remarkably similar evidence has recently been obtained in my group. As for faces, behavioural adaptation after-effects induced by “anti-voice” adapters induce greater perceptual shifts in speaker identification than non-opposite adapters (Latinus & Belin, 2011). Moreover, voices that are more acoustically different from an internal voice prototype (approximated by the morphing generated average of many voices) are perceived as more distinctive, and elicit greater activity in the TVAs, than voices

with a shorter distance to mean, more acoustically similar to the prototype (Latinus, McAleer, Bestelmeyer, & Belin, 2013) (Figure 2).

3. Social perception

Another aspect of face and voice processing that features baffling similarities is the formation of social percepts and inferences – how we judge someone unknown as attractive, competent or untrustworthy based on a glance at their face or a word heard.

- a. *Attractiveness and averaging*. It has long been established that averaging faces makes them more attractive – face composites generated by averaging faces from several different identities are judged more attractive than individual faces (Galton, 1878; Langlois & Roggman, 1990) – although it is clear that there is more to attractiveness than mere averageness (DeBruine, Jones, Unger, Little, & Feinberg, 2007). Two main explanations have been proposed for this phenomenon: the “good genes” account, proposing that we perceive averaged faces as more attractive because if they were real faces they would belong to individuals with high genetic fitness (with more symmetrical features, fewer imperfections), good potential mates (Grammer, Fink, Moller, & Thornhill, 2003; Langlois & Roggman, 1990; Thornhill & Gangestad, 1999); the “perceptual fluency” account, that averaged faces are preferred because they are closer in face space (more perceptually similar) to a putative internal prototype and hence easier to process (Halberstadt & Rhodes, 2003; Winkielman, Halberstadt, Fazendeiro, & Catty, 2006). Both accounts predict that a similar phenomenon should be observed for voices: that averaging the voice of several individuals should result in a more attractive voice.

Thanks to the recent advent of voice morphing tools (Kawahara & Matsui, 2003), we were able to test this prediction and indeed a steady increase of attractiveness ratings along with number of averaged voices in a composite was observed (Bruckert et al., 2010), similar that that found for faces (Figure 3). More than a curiosity, the phenomenon offers a window onto the perceptual mechanisms of voice attractiveness and reveals that two main, independent, acoustical

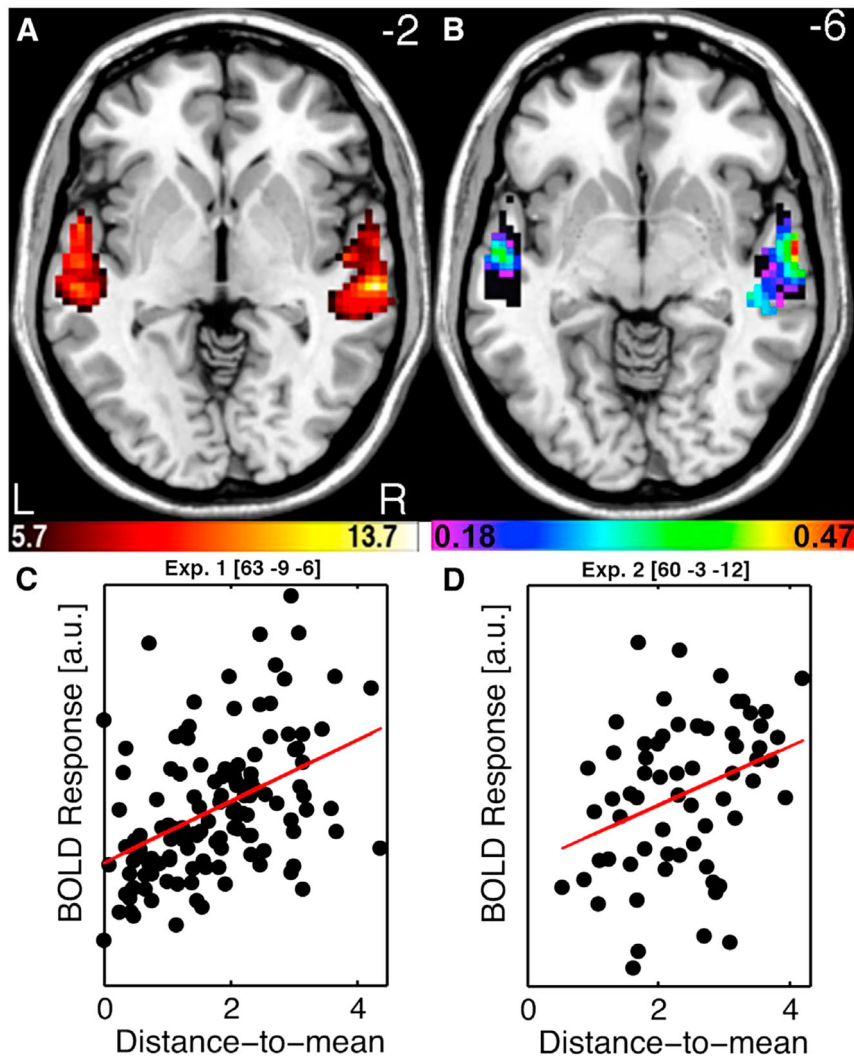


Figure 2. Norm-based coding of speaker identity in the Temporal Voice Areas. (A) TVA showing significantly greater fMRI signal in response to vocal versus nonvocal sounds at the group-level used as a mask for further analysis. Colour scale indicates T values of the vocal versus nonvocal contrast. (B) Maps of Spearman correlation between beta estimates of BOLD signal in response to each voice stimulus and its distance-to-mean overlay on the TVA map (black). Colour scale indicates significant r values ($p < .05$ corrected for multiple comparisons). Note a bilateral distribution with a maximum along the right anterior STS. (C) Scatterplots and regression lines between estimates of BOLD signal and distance-to-mean at the peak voxel for “had” syllables. (D) Scatterplots and regression lines between estimates of BOLD signal and distance-to-mean at the peak voxel observed for “hellos”. Reproduced (permission pending) from Latinus et al. (2013).

features are at play: the amount of spectro-temporal irregularities (measured for instance by the harmonics-to-noise ratio) and “distance-to-mean” where voices acoustically more similar to the average are generally found more attractive (Bruckert et al., 2010). These two acoustical parameters are directly analogous to two features known to be key determinants of facial attractiveness: face texture smoothness and distance to mean.

- b. *Social inferences in a same 2D “social space”.* Social face perception research has shown that people

readily form personality impressions from unknown faces. These social inferences are formed rapidly – a mere second is needed to reach competence judgments that predict election margins (Todorov, Mandisodza, Goren, & Hall, 2005) – and robustly: impressions may not be accurate but different people tend to agree on them. The diverse personality impressions – competence, aggressiveness, friendliness – are well summarized by a 2D “social face space” with perceived Trustworthiness and Dominance as the two main axes (Oosterhof & Todorov,

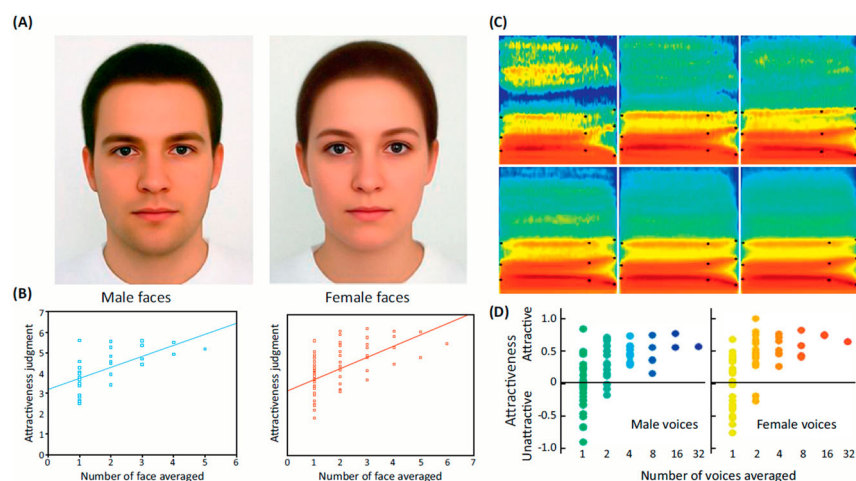


Figure 3. Averaging and attractiveness in faces and voices. Face and voice attractiveness judgments as a function of averaging. (A) Face composites generated by averaging 32 male faces (left) and 64 female faces (right). (B) Attractiveness ratings as a function of number of face averaged. Note the steady increase in attractiveness ratings with increasing number of averaged faces, for both male (left) and female (right) faces. Reproduced, with permission, from Braun, Gruendl, Marberger, and Scherber (2001). (C) Spectrograms of voice composites generated by averaging an increasing number of voices of the same gender (different speakers uttering the syllable “had”). (D) Attractiveness ratings as a function of number of voices averaged. Note the steady increase in attractiveness ratings with increasing number of averaged voices, for both male (left) and female (right) voices. Reproduced (permission pending) from Bruckert et al. (2010).

2008). We recently found that the exact same conclusions can be reached for voices. From a single word “hello”, listeners were found to form personality impressions with high inter-rater agreement, and principal component analysis reveals that the diverse personality judgments are also well summarized in a 2D “social voice space” with main axes best corresponding to Trustworthiness and Dominance, exactly as for faces – a finding observed for both male and female voices (McAleer, Todorov, & Belin, 2014).

The above sections have provided a short and voluntarily partial, but I believe no less striking, account of the similarities in face and voice processing. These similarities, observed in several different domains – discrimination from other stimuli, processing of identity information, social inferences – are perhaps not so surprising, considering that the processing of these different types of information poses similar problems to the cortical architecture, that are apparently implemented using similar solutions; which also no doubt facilitates the integration of this information across different modalities during naturalistic encounters (Campanella & Belin, 2007).

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Biotechnology and Biological Sciences Research Council [Grant Number BB/E003958/1]; Economic and Social Research Council [Grant Number RES-060-25-0010]; Fondation pour la Recherche Médicale [Grant Number AJE2012-14] and Labex BLRI [ANR-11-LABX-0036] supported by the French National Agency for Research (ANR) under the program “Investissements d’Avenir” [ANR-11-IDEX-0001-02].

ORCID

Pascal Belin  <http://orcid.org/0000-0002-7578-6365>

References

- 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. *Current Biology*, *24*(5), 574–578. doi:10.1016/j.cub.2014.01.058
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker’s voice in right anterior temporal lobe. *Neuroreport*, *14*(16), 2105–2109. doi:10.1097/00001756-200311140-00019
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*, 17–26
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565

- Bestelmeyer, P., Belin, P., & Grosbras, M. H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, 21, R838–R839.
- Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience & Biobehavioral Reviews*, 47, 717–734. doi:10.1016/j.neubiorev.2014.10.022
- Bodamer, J. (1947). Die prosop-agnosie. *Archiv fur Psychiatrie und Nervenkrankheiten Vereinigt mit Zeitschrift fur die Gesamte Neurologie und Psychiatrie*, 179, 6–53.
- Braun, C., Gruendl, M., Marberger, C., & Scherber, C. (2001). Beautycheck - Ursachen und Folgen von Attraktivitae. *Neuroscience & Biobehavioral Reviews*, 25, 101–110.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305–327.
- Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., ... Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116–120. doi:10.1016/j.cub.2009.11.034
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6, 641–651.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11, 535–543.
- Capilla, A., Belin, P., & Gross, J. (2013). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. *Cerebral Cortex*, 23(6), 1388–1395. doi:10.1093/cercor/bhs119
- Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S., ... Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, 10, 127. doi:10.1186/1471-2202-10-127
- DeBruine, L. M., Jones, B. C., Unger, L., Little, A. C., & Feinberg, D. R. (2007). Dissociating averageness and attractiveness: Attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1420–1430. doi:10.1037/0096-1523.33.6.1420
- Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, 16, 166–173.
- Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005), 845–851. doi:10.1126/science.1194908
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12, 1187–1196.
- Galton, F. (1878). Composite portraits. *Journal of the Anthropological Institute*, 8, 132–144.
- Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., ... Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia*, 47, 123–131. doi:10.1016/j.neuropsychologia.2008.08.003
- Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the National Academy of Sciences*, 101, 17516–17521.
- 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. *Nature Neuroscience*, 9(8), 1064–1070.
- Grammer, K., Fink, B., Moller, A. P., & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews*, 78(3), 385–407.
- Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin & Review*, 10(1), 149–156.
- Haxby, J. V., Hoffman, E. A., & Ida Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4, 223–233.
- 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(3), 1376–1389. doi:10.1016/j.neuroimage.2012.05.070
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neuroscience*, 17(11), 4302–4311.
- Kawahara, H., & Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 256–259.
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces Are only average. *Psychological Science*, 1(2), 115–121.
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2, 175. doi:10.3389/fpsyg.2011.00175
- Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, 23(12), 1075–1080. doi:10.1016/j.cub.2013.04.055
- Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102), 572–575.
- Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4, 89–94.
- Linden, D. E., Thornton, K., Kuswanto, C. N., Johnston, S. v. d. V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, 21, 330–337.
- Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, 8, 1386–1391.
- McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLoS One*, 9(3), e90779.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. doi:10.1073/pnas.0805664105

- Ortiz-Rios, M., Kusmieriek, P., DeWitt, I., Archakov, D., Azevedo, F. A., Sams, M., ... Rauschecker, J. P. (2015). Functional MRI of the vocalization-processing network in the macaque brain. *Frontiers in Neuroscience, 9*, 14608. doi:10.3389/fnins.2015.00113
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., ... Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NEUROIMAGE, 119*, 164–174. doi:10.1016/j.neuroimage.2015.06.050
- 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 in Cognitive Sciences, 19*(12), 783–796. doi:10.1016/j.tics.2015.09.002
- Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Current Biology. doi:10.1016/j.cub.2011.07.028*
- Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience, 11*(2), 367–374.
- Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology, 19*(4), 319–324.
- Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*, 1568–1573.
- Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46*(18), 2977–2987.
- Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience, 6*, 258–307.
- Roswandowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology, 24*(19), 2348–2353. doi:10.1016/j.cub.2014.08.048
- Sergent, J., & Signoret, J. L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex, 2*, 375–388.
- Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452–460.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. doi:10.1126/science.1110589
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience, 6*, 989–995.
- 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.
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437. doi:10.1146/annurev.neuro.30.051606.094238
- 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.
- Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical & Experimental Neuropsychology, 11*, 665–674.
- 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.
- Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science, 17*(9), 799–806.
- Xu, X., Biederman, I., Shilowich, B. E., Herald, S. B., Amir, O., & Allen, N. E. (2015). Developmental phonagnosia: Neural correlates and a behavioral marker. *Brain and Language, 149*, 106–117. doi:10.1016/j.bandl.2015.06.007
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*(4), 959–974. doi:10.1111/j.2044-8295.2011.02045.x
- Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences, 17*(6), 263–271. doi:10.1016/j.tics.2013.04.004