

Voice processing in human and non-human primates

Pascal Belin^{1,2,*}

¹*Centre for Brain, Music and Sound (BRAMS), Département de Psychologie, Université de Montréal, CP 6128 Succ Centre-Ville, Montréal, Québec, H3C 3J7 Canada*

²*Centre for Cognitive Neuroimaging (CCNi), Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK*

Humans share with non-human primates a number of voice perception abilities of crucial importance in social interactions, such as the ability to identify a conspecific individual from its vocalizations. Speech perception is likely to have evolved in our ancestors on the basis of pre-existing neural mechanisms involved in extracting behaviourally relevant information from conspecific vocalizations (CVs). Studying the neural bases of voice perception in primates thus not only has the potential to shed light on cerebral mechanisms that may be—unlike those involved in speech perception—directly homologous between species, but also has direct implications for our understanding of how speech appeared in humans. In this comparative review, we focus on behavioural and neurobiological evidence relative to two issues central to voice perception in human and non-human primates: (i) are CVs ‘special’, i.e. are they analysed using dedicated cerebral mechanisms not used for other sound categories, and (ii) to what extent and using what neural mechanisms do primates identify conspecific individuals from their vocalizations?

Keywords: voice; speech; conspecific vocalizations; speaker recognition; auditory cortex; social cognition

1. INTRODUCTION

In the auditory environment of primates, vocalizations produced by a conspecific individual—conspecific vocalizations (CVs)—are sounds of overriding importance. Most non-human primates possess a rich vocal repertoire, which they use in many different contexts, such as agonistic or affiliative interactions with members of their social group, territorial calls and alarm calls, many of them loud enough to be heard at a distance (e.g. Winter *et al.* 1966; Green 1975). Thus, each individual is daily exposed to a large number of CVs from several callers (Snowdon 1986; Hauser 1996). In humans, particularly in modern societies, voices are everywhere, from physically present individuals as well as increasingly from virtual sources such as radios, TVs, etc., and we spend a large part of our time listening to these voices.

CVs are extremely rich in information. The clearest example is human speech, a uniquely human adaptation to transmit symbolic information in a highly efficient manner—although precursors of speech may exist in non-human primates as well (Seyfarth *et al.* 1980; Hauser *et al.* 2002). Speech played a major role in our global domination of other species. Accordingly, much research effort in auditory perception has focused on speech perception. Yet, speech perception constitutes the tip of an iceberg of pre-existing cognitive abilities to extract information contained in CVs.

Primate vocalizations—human as well as non-human—can be thought of as ‘auditory faces’ (Belin *et al.* 2004) that carry in their acoustic structure a

wealth of paralinguistic information. As our face, the human voice carries much information on our physical characteristics and our affective state; for example, allowing recognition of a person over the telephone. Accurate perception of this information plays a major role in our social interactions. Similarly, the cooperative structure and frequent social interactions of most non-human primates emphasize the importance of good abilities to accurately extract information in CVs. Examples of increased chances of mating success and survival related to accurate perception of vocal information include: accurate perception and appropriate response to predators’ alarm calls; rapid recognition by a mother that the distress calls she hears are from her infant; accurate evaluation of reproductive fitness in the call of a potential mate during courtship, etc.

The nervous system of our primate ancestors has therefore been subject to high evolutionary pressure to develop neural mechanisms endowing primates with abilities to rapidly and accurately categorize relevant information in CVs, turning them into ‘auditory specialists’ (Ghazanfar & Santos 2003). Many of these ‘voice perception’ abilities are probably shared to a large extent between human and non-human primates—unlike speech perception. Our understanding of the communicative brain can only be increased by a closer study of vocal cognitive abilities having emerged in all primates as similar solutions to common ecological problems, perhaps based on similar cerebral mechanisms as well.

This review adopts a comparative perspective to examine behavioural and neurobiological evidence relative to two main questions that can be posed in

*p.belin@psy.gla.ac.uk

One contribution of 14 to a theme issue ‘The neurobiology of social recognition, attraction and bonding’.

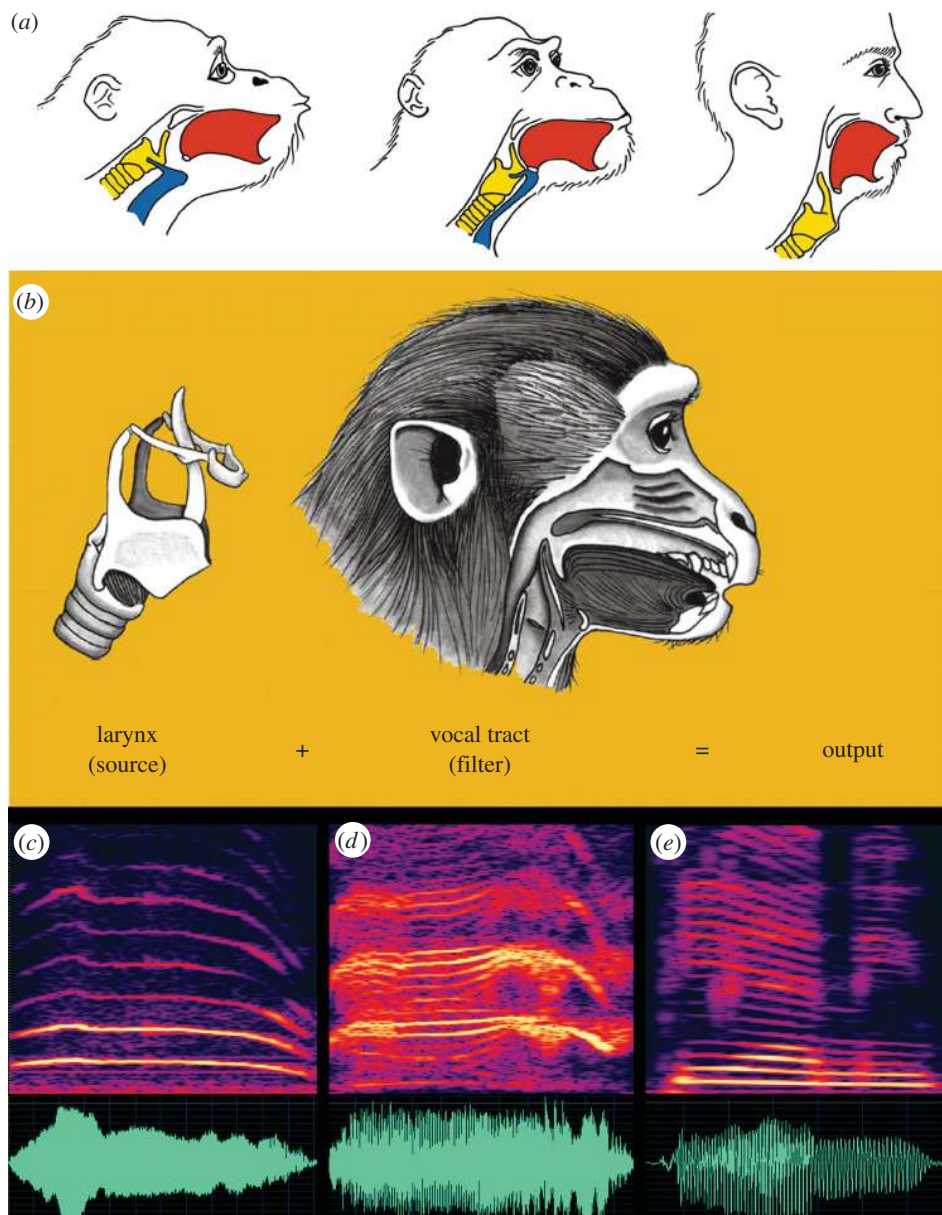


Figure 1. Voice production mechanism in primates. (a) Sagittal views depicting vocal tract anatomy in an (i) orang-utan, (ii) a chimpanzee and (iii) a human. Red colour, the tongue body; yellow, the larynx; blue, the air sacs (apes only). Note the longer oral cavity and much lower larynx in the humans, with concomitant distortion of tongue shape compared with orang-utans and chimpanzees. These differences allow a much greater range of sounds to be produced by humans, which would have been significant in the evolution of speech (Fitch 2000). Adapted with permission from Fitch (2000). (b) The source/filter theory. The source/filter theory of vocal production, originally proposed for speech, appears to apply to vocal production in all mammals studied so far. The theory holds that vocalizations result from a sound source (typically produced at the larynx) combined with a vocal tract filter (which consists of a number of formants). This filtering action applies regardless of the type(s) of sound produced at the larynx. Reproduced with permission from Fitch (2000). (c)–(e) Spectrograms (0–5500 Hz) of a rhesus coo (c), a chimp pan-hoot excerpt (d) and human speech (e). Note the similarities in structure, with harmonics and formants visible in each case.

similar terms for human and non-human primates. The first question is whether the perception of CVs involves specific neuronal processes or not compared with non-vocal sounds or heterospecific vocalizations. In other words, are CVs 'special'? The second question concerns the ability to extract identity information in voices. Can our non-human relatives recognize callers by their vocalizations? What are the neuronal correlates of these voice recognition abilities?

Before addressing these two issues, we will begin with a rapid overview of similarities and differences in the voice production mechanisms of human and non-human primates.

2. HUMAN VOICE AND PRIMATE VOCALIZATIONS

The human voice and non-human primate vocalizations share a number of similarities in their acoustic structure, but are also characterized by important differences (Fitch 2000, 2003). The basic mechanism of voice production is similar across primates (figure 1); the 'source/filter theory' developed by Fant (1960) in the context of human speech production (Fant 1960) also largely applies to non-human primate vocalizations (Owren & Linker 1995; Fitch 2000). Briefly, the sound source produced in the larynx generally consists of quasi-periodic series of pulses generated by the successive openings and closings of the vocal folds

(the 'mucosal wave'; Titze 1993). The rate of vibration of the vocal folds determines the fundamental frequency of phonation (f_0). The frequency spectrum of this sound source contains energy not only at the f_0 , but also at all integer multiples of the f_0 (harmonics). In addition to this quasi-periodic component, the source contains some proportion of inharmonicity, such as temporal irregularities in vocal fold vibration (contributing to the 'rough' quality of voice) or noise caused by aerodynamic turbulences (contributing to the 'breathy' quality of voice; the source exclusively consists of turbulent noise in the case of whispered speech). Besides the 'modal' register described above, humans as well as monkeys and apes are also able to use the larynx in different modes with varying degrees of nonlinearity, such as the 'falsetto' and the 'vocal fry' registers in humans (Eskenzi *et al.* 1990).

The sound emitted by the larynx is modified by the cavities and tissues located above the larynx (supralaryngeal vocal tract), which act as an acoustic filter relatively independent of the source characteristics—for example, unlike in wind instruments (Fant 1960; Fitch 2003). The vocal tract causes resonances—the 'formants'—that reinforce energy at certain frequencies depending on the shape of the vocal tract (figure 1). In humans, different vowels correspond to different configurations of the articulators that yield different resonant properties of the vocal tract, and thus induce formants at different frequencies. These formant frequencies constitute a critical acoustic cue for the identification of vowels: speech synthesizers based on formant synthesis achieve a high degree of realism in vowel production using a single source and as little as three formants (e.g. Klatt 1980); reasonable rates of speech recognition can even be obtained from sine-wave analogues of speech composed only of three pure tones following the frequencies of the first three formants (Remez *et al.* 1981). In non-human primates, there is now ample evidence in several species that the laryngeal sound source is also subject to spectral patterning by the supralaryngeal vocal tract, resulting in formants clearly visible on sonograms (Owren *et al.* 1997; figure 1). Thus, monkeys and apes use vocalizations that combine source due to vocal fold movement with filtering by vocal tract—comparable to our vowels (Rendall *et al.* 1998). For example, baboon grunts are very similar to our neutral, central vowel pronounced with a relaxed vocal tract (Owren *et al.* 1997).

Apart from these similarities, the human voice differs from non-human vocalizations on several important aspects. There are a number of morphological differences in the vocal apparatus of human and non-human primates (Fitch 2000, 2003). Comparative studies of the anatomy of the vocal folds show that several species of monkeys possess vocal membranes (or vocal lips), consisting of thin extensions of the vocal folds lacking in humans (Schön Ybarra 1995). Although the issue needs further investigation, the very low mass of these vocal membranes is thought to enable the production of vocalizations with high pitch (Fitch 2003). Another difference is that many primates possess air sacs in the larynx—out-pouchings of the epithelium lining the larynx—not present in humans. The exact role of these sacs is still unclear, although

they are thought to be involved in the loud calls of some species (Schön Ybarra 1995; Fitch 2003).

A more important difference in vocal tract anatomy between human and non-human primates concerns the position of the larynx; the human larynx is much lower than its non-human counterpart (except in very young infants). This particularity has the rather harmful consequence of forcing particles of food or water to pass in front of the trachea before reaching the entrance of the oesophagus, at significant risk of entering into the lungs—causing hundreds of accidental deaths every year. The exact nature of the evolutionary advantage provided by this 'descent of the larynx' is still debated, but it must have been quite important to compensate for the associated significant increase in risk of choking. One clear advantage conferred by the descended larynx is the increased space for the tongue, which as a consequence is much less elongated and more flexible in humans (Schön Ybarra 1995). Another possible advantage is that a descended larynx directly lengthens the vocal tract, thus lowering formant frequencies. Since vocal tract length is generally well correlated with body size, the descended larynx could contribute to convey an exaggerated perceptual impression of size in listeners (Fitch 2000, 2003).

Especially important consequences of the lowered larynx in humans are an increased flexibility of the tongue and an important angle in the vocal tract, both yielding an increased range of variation in formant frequencies. The typical 'vowel space' of non-human primates is smaller (corresponding to less formant variations) due to the relatively inflexible nature of their vocal tract (Lieberman *et al.* 1969). Thus, non-human primates have a lesser ability to create several acoustically distinctive sounds from a same source through supralaryngeal vocal tract filtering.

In sum, the human vocal tract is characterized by several morphological differences that probably contributed to/accompanied the emergence of speech. Yet, the basic mechanisms of voice production are largely similar between humans and our non-human relatives, yielding similar acoustic structures (figure 1) and comparable influence of inter- and within-individual variability. This in turn posed similar ecological problems to the brain of the receiver, which may have been solved using similar cerebral mechanisms across species of primates.

3. ARE CONSPECIFIC VOCALIZATIONS SPECIAL?

One essential question relative to the cerebral organization underlying voice perception abilities is whether these neural mechanisms are exclusively dedicated to process CVs or are also involved analysing other classes of sounds. In other words: 'are primate CVs special?'—or in the case of humans, 'are voices special?' (This question has been asked many times in the domain of face perception: the 'are faces special?' question still generates much argument and research; Farah 1996; Kanwisher *et al.* 1997; Gauthier *et al.* 2000; Haxby *et al.* 2001). In this section, we review behavioural and neurobiological evidence for species-specific mechanisms in the perception of vocalizations, in non-human primates as well as in humans.

(a) Behavioural evidence in non-human primates

It is clear from observing the behaviour of non-human primates that they are particularly influenced by hearing CVs. One relevant question is whether the special status of these sounds is associated with enhanced measures of perceptual sensitivity for discriminating CVs. Zoloth *et al.* (1979) used an operant paradigm with food rewards to train several species of Old World monkeys to discriminate between variants of 'coo' calls from Japanese macaques (*Macaca fuscata*). The discrimination could be based either on the temporal position of the f₀ peak in the call ('smooth early', SE, versus 'smooth late', SL), an acoustic cue of behavioural relevance for Japanese macaques, since it distinguishes between variants used in different contexts, or on the starting pitch of the call, an acoustic cue with no particular relevance. Japanese macaques were found to perform much better than the comparison species when the discrimination task was based on the behaviourally relevant temporal cue; in contrast, they were worse than the other monkeys when the discrimination was based on the irrelevant dimension and pitch (Zoloth *et al.* 1979). Thus, this study provides strong evidence in one species of Old World monkeys for an enhanced perceptual discrimination of CVs compared with other species; however, this seems to hold only if the discrimination is based on an ecologically valid contrast.

A connected question is how do non-human primates perceive human speech sounds. Many studies have used speech material to probe auditory perceptual abilities of non-human primates. One study compared difference limens of humans and monkeys at a discrimination task using synthetic consonant–vowel English syllables. The syllables were arranged in a continuum of voice onset time (VOT), an important cue for a place of articulation (Sinnott & Adams 1987). Humans were found to discriminate pairs of syllables with differences in VOT two to four times smaller than the monkeys. Sinnott (1989) also found that monkeys were less accurate than human listeners to discriminate synthetic English vowels. However, the pairs with which the monkeys had most difficulties were also the ones that led to the longest reaction times in humans, suggesting comparable analysis mechanisms (Sinnott 1989). Ramus *et al.* (2000) found that both cotton-top tamarins as well as human babies were able to discriminate speech sentences in two different languages, but not if the sentences were played backwards (Ramus *et al.* 2000). Conversely, Hopp *et al.* (1992) found that Japanese macaques were less accurate than human listeners in a discrimination task along a continuum of synthetic 'coos' varying on the temporal position of the f₀ peak—although humans also generally perform better at discrimination tasks involving lower-level acoustic cues (Owren *et al.* 1992).

(b) Neurobiological evidence in non-human primates

Does the nervous system of non-human primates show a specialization for processing CVs? One possible sign of species specificity in the processing of vocalizations is that primates seem to have increased sensitivity at

frequencies corresponding to the range found in their species-specific vocalizations (Aitkin *et al.* 1986; Wang 2000). For example, monkeys have better sensitivity (smaller absolute auditory thresholds) than humans in high, but not in low frequencies (Owren *et al.* 1988), consistent with the higher frequency range of monkey vocalizations compared with human voice. More evidence is needed to allow the generalization of this observation to all primates.

Several teams have used electrophysiological recordings in awake non-human primates to investigate the response of auditory cortex to various sound categories, including conspecific calls. One of the first set of studies was performed in the squirrel monkey, a highly vocal New World primate whose vocal behaviour is well documented (Newman 2003). Winter & Funkenstein (1973) found that more units responded to pure tones than to conspecific calls in auditory cortex, but for the first time evidenced a small number of cells that responded only to CVs. In a subsequent study using a larger set of conspecific calls, they found that more than half the cells that responded to CVs displayed some selectivity in responding only to no more than two acoustically similar calls (Winter & Funkenstein 1973). These results initially suggested the existence of 'call detectors', i.e. specialized neurons responding only to CVs. However, subsequent experiments using more repetitions of the same calls found that vocalization-responsive neurons of auditory cortex typically responded to more than one call or to various features of calls (Wollberg & Newman 1972); moreover, their response properties were in fact quite variable and were found to change significantly over the course of an hour (Manley & Muller-Preuss 1978). More recently, Wang *et al.* (1995) suggested that cells in the primary auditory cortex (A1) of the marmoset could be categorized into two general classes: one responding to call types and another to a wider range of sounds, including vocalizations as well as non-vocal sounds (Wang *et al.* 1995).

Thus, at least at the primary stages of auditory cortex, CVs typically elicit strong responses in a large proportion of cells; however, the notion of 'call detectors' or neurons highly specialized for processing CVs now seems doubtful, and this is progressively replaced by the idea of population coding where features of the vocal signal are coded by the distributed activity of a large number of cells (Wang 2000; Newman 2003).

One way to better characterize the specificity of response to CVs is to compare the cellular responses to CVs and time-reversed versions of the same calls, i.e. stimuli with the same spectral structure but a different temporal structure and lacking the natural behavioural meaning of these calls. Glass & Wollberg (1983) found in the awake squirrel monkey that the responsiveness of cells of both primary and secondary auditory cortices was not significantly different from calls or their time-reversed versions; very few cells were found to show 'reversed responses' to the time-reversed vocalizations (Glass & Wollberg 1983). However, a more recent study in the anaesthetized common marmoset found that a majority of A1 neurons showed stronger responses to natural marmoset twitter calls than to

their time-reversed version (Wang *et al.* 1995). While this finding is consistent with similar recent findings in other species, it may simply reflect the lack of naturalness of the reversed calls and not species specificity *per se* in the processing of CVs; other stimuli such as heterospecific vocalizations or other natural sounds might very well yield the same result.

Some of the strongest evidence against this alternative explanation has been obtained by Wang & Kadia (2001), who compared the responses elicited by natural and time-reversed marmoset twitter calls in A1 neurons of the cat, a species for which neither the natural nor the reversed version of the twitter call have ecological relevance. They found that contrary to marmoset neurons, cat A1 neurons did not respond differently to the natural and time-reversed versions of the call (whereas they were found to do so for cat vocalizations; Wang & Kadia 2001). Moreover, this lack of preference appeared to be due to weaker responses to the natural calls in the cat than in the marmoset A1, whereas responses to the time-reversed calls were comparable in the two species (Wang & Kadia 2001). The diminished response of A1 neurons to time-reversed calls in marmosets is thus not only related to their lack of naturalness, as it was not observed in another species for which the time-reversed call would be presumably as unnatural. However, the stronger response to natural than time-reversed stimuli is not necessarily the signature of a species-specific mechanism, as it could also be related to the ecological value of the call; behaviourally relevant sounds from other species (such as predators or humans) might also induce a similar pattern of response. A stronger test of species specificity in the processing of CVs thus will ultimately require comparison of CVs with a larger array of natural sounds.

Thus, there is no clear demonstration yet of neuronal mechanisms selectively engaged by CVs in a non-human primate A1. What is the present evidence in other parts of auditory cortex? Tian *et al.* (2001) recorded from neurons in the lateral belt of lightly anaesthetized rhesus monkeys in response to the presentation of seven conspecific calls presented at seven azimuthal locations (Tian *et al.* 2001). In all three regions of the lateral belt (anterolateral, AL; medio-lateral, ML; caudolateral, CL), neurons were found to display some call selectivity (i.e. more than half of the cells responding with more than 50% of their maximal firing rate to three calls or less out of the seven calls; Tian *et al.* 2001). In particular, selectivity was found to be significantly better in the AL field, which the authors interpreted as evidence for a 'what' (object identification) versus 'where' (spatial localization) functional segregation between anterior and posterior fields, as in primate visual cortex (Ungerleider & Haxby 1994; Kaas & Hackett 1999; Rauschecker & Tian 2000).

Neurons responding to sounds and, in particular, CVs have also been observed outside auditory cortex. Romanski & Goldman-Rakic (2002) identified what seems to constitute an auditory responsive region in the prefrontal cortex of awake-behaving rhesus macaques. Neurons in a discrete region of ventrolateral prefrontal cortex were found to respond to complex sounds, including CVs and human vocalizations. Most neurons in this auditory domain responded to both vocalizations

and non-vocalization stimuli, but most ($n=52/70$) responded more strongly to vocalizations, and a small subset of cells ($n=3$) responded only to macaque or human vocalizations, with at least one cell responding only to a CV (Romanski & Goldman-Rakic 2002). More recently, Romanski *et al.* (2004) investigated in greater detail the response of these auditory prefrontal cells to CVs, using a large set of CVs from several different callers. The majority of the recorded cells was found to respond to between two and five vocalizations, with 2/301 cells being caller-selective. However, because this last study focused on the response to CVs, cells responsive to CVs were not tested with non-vocalization stimuli, not allowing any conclusion to be drawn on the possible vocalization specificity of these cells.

In sum, a large body of electrophysiological studies in non-human primates has evidenced many cells with significant responses to CVs in primary and secondary auditory cortices, as well as in ventrolateral prefrontal cortex. Yet, few studies to date have systematically compared responses elicited by CVs with those elicited by heterospecific vocalizations or by equally complex, non-vocalization stimuli. Only one study so far has reported cells that seemed to respond only to CVs (Romanski & Goldman-Rakic 2002), although in very small proportion (one or two on 400 recorded cells). Thus, it seems too early to conclude unequivocally on the species specificity of the mechanisms involved in processing CVs in non-human primates.

(c) *Functional lateralization in processing CVs*

Several studies have measured indexes of functional lateralization in the processing of CVs by non-human primates seeking to demonstrate an advantage of the left hemisphere. The rationale behind these studies is that left-lateralized processing of CVs in non-human primates might provide an evolutionary precursor of the left-hemisphere advantage for speech processing in humans.

Petersen *et al.* (1978) used a psychophysical paradigm similar to the one used by Zoloth and colleagues (see §3a) to train several species of Old World monkeys to discriminate between variants of 'coo' calls from Japanese macaques (*M. fuscata*), except that stimuli were presented monaurally either to the left or to the right ear. When the discrimination was based on the communicatively relevant peak position (SE versus SL), they found that all five Japanese macaques they had trained made less errors when CVs were presented to the right as compared with the left ear; such a right-ear advantage was only observed in one out of five of the animals from comparison species. In contrast, Japanese macaques trained to perform the discrimination based on the pitch dimension showed either a left-ear advantage or no advantage. A follow-up study by the same group replicated the findings of a right-ear advantage in Japanese macaques in the discrimination of SE versus SL versions of their coos, and further confirmed that comparison animals failed to show lateralized processing although they were using similar acoustic dimension in their judgement (Petersen *et al.* 1984).

These results provide strong evidence that left-lateralized neural mechanisms analogous to those observed in human speech processing can be engaged

in Japanese macaques when they attend selectively to the temporal position of the f0 peak in their coo (SE versus SL). The lack of lateralization in comparison animals in two consecutive studies is particularly interesting, since it suggests that these lateralized processes could be observed only for conspecific calls; yet a complete verification of this hypothesis would have required animals to be tested with vocalizations from the comparison species as well. The fact that Petersen *et al.* (1978) did not observe a right-ear advantage when the same sounds were discriminated by pitch—although only in two animals—could be interpreted along with the authors as suggesting that only the communicatively relevant features of the call might engage lateralized processes. Alternatively, different features of the same call could be processed using partially distinct, differentially lateralized neural networks, as seems to be the case in humans; speech processing engages left-lateralized networks in most right-handed human subjects, but processing of pitch or identity from the same vocal input reverses this pattern and yields a right-hemisphere advantage (Zatorre *et al.* 1992; von Kriegstein *et al.* 2003).

Another method used to measure functional asymmetries in non-human primates involves unilateral cortical lesions. Heffner & Heffner (1984) used a variant of the paradigm used by Petersen *et al.* (1984) to train Japanese macaques to perform the discrimination of SE versus SL among 15 different coos. Then they performed unilateral lesions in the superior temporal gyrus encompassing primary as well as secondary auditory cortices and measured the effects of the lesions on performance at the coo discrimination according to whether the lesion had been performed in the left (five animals) or in the right (five animals) hemisphere. A striking pattern of lateralization emerged: the animals having received a lesion in the right hemisphere showed no noticeable deficit when tested within 3–8 days of the lesion, whereas the animals with a lesion in the left hemisphere showed a marked initial deficit followed by a progressive recovery over the following days. A second lesion to the remaining auditory cortex of the other hemisphere then completely abolished the ability to discriminate the coos (Heffner & Heffner 1984). The monkeys were still able to perform simpler discrimination of coos from noise or tones outside the frequency range of coos (2 and 4 kHz; Heffner & Heffner 1986). Thus, these results are consistent with the findings of Petersen *et al.* (1978, 1984) in suggesting that the discrimination of SE and SL coos primarily engages the left hemisphere in Japanese macaques.

Playback experiment in field studies has also yielded useful information on the cerebral lateralization of the processing of CVs. Hauser & Andersson (1994) monitored the orienting response to CVs in a large number of free-ranging rhesus macaques in the colony of Cayo Santiago. The sounds were played exactly 180° behind the experimental animal while feeding on one of the three food dispensers of the island, so that the target animal could choose to orient to the source by turning the head either to the left or to the right. The majority of adult macaques (61 out of 80) was found to orient to the sound source by turning their head to the right, thus

seeking to increase sound amplitude in the right ear, or the left hemisphere, whereas they tended to present the left ear to the source when a familiar, but heterospecific alarm call was played. Infants tested using the same paradigm failed to show any head-turning preference. The authors interpreted this finding as evidence for left-biased cerebral lateralization for processing CVs in the rhesus macaque, as for human speech, but only once a certain stage of maturation is reached. Follow-up studies using the same paradigm but acoustically modified CVs replicated the right-ear orienting bias in the adult rhesus monkeys, and further showed that temporal modifications such as expansion of contraction (Hauser *et al.* 1998) or temporal inversion (Ghazanfar *et al.* 2001) could eliminate or reverse the right-ear advantage.

(d) *Neuroimaging studies in non-human primates*

More recently, several teams used neuroimaging techniques generally used in humans to measure cerebral activity during processing of CVs in awake monkeys. Poremba *et al.* (2004) used positron emission tomography (PET) to measure metabolic activity in rhesus macaques during passive listening to several classes of complex sounds, including CVs, human vocalizations, as well as non-vocal sounds from the environment (Poremba *et al.* 2004). Each superior temporal gyrus was divided into five regions of interest, and metabolic activity in each region was compared across hemispheres. Unexpectedly, all sound categories elicited stronger activity in the right than in the left hemisphere in the posterior parts of the superior temporal gyrus corresponding to auditory cortex. Yet, a left-lateralized pattern of activity was found in the dorsal temporal pole, the most anterior region of interest, only for the conditions where CVs were present (CVs or CVs mixed with other sounds). These findings were interpreted as suggesting that the temporal pole might constitute a precursor of a human acoustic language area (Poremba *et al.* 2004). Of particular interest would have been a comparison of activity across the different classes of sounds within a same region. This comparison, unfortunately not provided, would have had the potential to uncover possible regions of specific response to CVs in non-human primates.

Gil-da-Costa *et al.* (2004) used PET in awake macaques to measure cerebral blood flow during auditory stimulation with CVs (coos and screams) and non-biological sounds. They found that CVs elicited greater activity than non-biological sounds in several posterior visual-processing regions extending from early to higher-order areas in the ventral object-processing stream and in visual motion-processing areas extending to posterior superior temporal sulcus (STS; Gil-da-Costa *et al.* 2004). Interestingly, they also found CVs to elicit greater activation than the non-biological sounds in several peri-sylvian areas, including area Tpt in the posterior superior temporal gyrus, as well as in the ventrolateral portions of the STS (Ricardo Gil-da-Costa, personal communication). Unfortunately, however, neuronal activity was not measured during stimulation with intermediate control categories, such as biological sounds or heterospecific vocalizations. Thus, the species

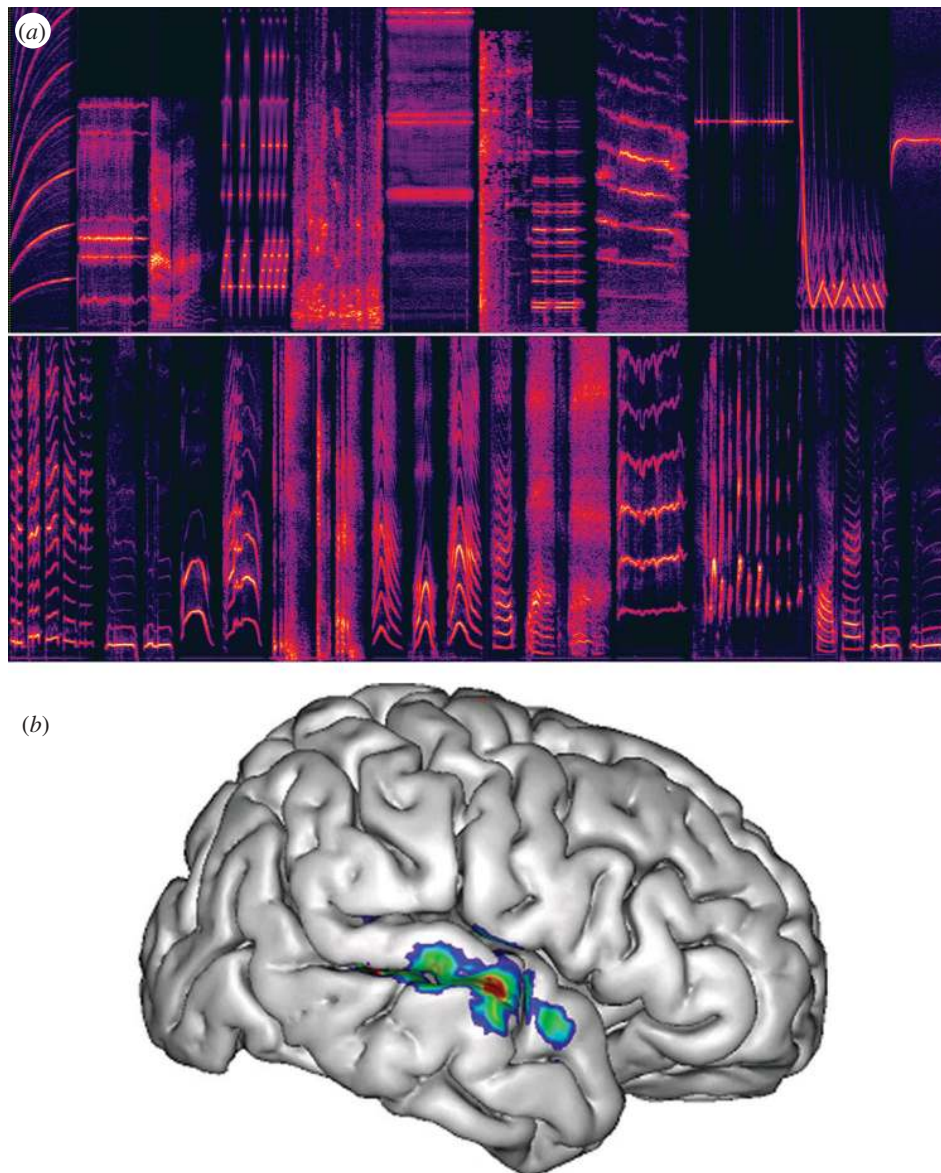


Figure 2. STS voice-selective areas in humans. (a) Spectrograms (0–5000 Hz) of examples of (i) non-vocal and (ii) vocal sounds used by Belin *et al.* (2000). Note their similar apparent complexity. (b) Cortical rendering of regions showing greater response to vocal compared with non-vocal sounds in eight subjects, located in the anterior part of the STS. Reproduced with permission from Belin *et al.* (2004).

specificity of these activations remains to be demonstrated in future studies.

(e) *Specialization for voice perception in humans*

Humans have presumably been subject to similar evolutionary pressure as non-human primates to develop mechanisms specialized in accurately extracting information in CVs (voice). The paramount importance of speech in all human societies makes it even more probable that specific mechanisms have evolved in the human brain to process sounds of voice. What is the present evidence for such voice-selective mechanisms in humans?

Studies of patients with cerebral lesions constitute one crucial source of information. It is well known that lesions in the region of the left posterior superior temporal gyrus lead to the syndrome known as ‘Wernicke’s aphasia’, which is characterized among other things by a severe deficit in speech comprehension (Wernicke 1874; Damasio 1992). Another syndrome known as ‘pure

word deafness’, reported to occur after lesions involving the primary auditory cortex bilaterally (Shoumaker *et al.* 1977; Coslett *et al.* 1984), is characterized by a deficit that appears restricted to sounds of speech. In these two syndromes, the perception and recognition of other sounds such as music or sounds from the environment appear essentially preserved, which suggests that the deficits are restricted to human speech and makes a strong case for species specificity in humans’ auditory processing. However, this is not very surprising since speech is unique to humans.

Is there evidence for other acquired deficits restricted to human voice perception but not to speech? As noted by several authors, speech is but only one type of information contained in voice. The human voice contains a wealth of paralinguistic information, such as information on the speaker’s identity (gender, approximate age, etc.) and affective state, and a sound of voice may very well contain no speech at all (e.g. laughs, cries). These types of information are also present

to some extent in the vocalizations of non-human primates. Thus, evidence for human mechanisms selectively involved in extracting paralinguistic information in voice would be particularly useful, given our comparative perspective.

Such evidence exists and comes from the study of patients with deficits in voice discrimination or recognition, a deficit termed 'phonagnosia' (Van Lancker & Canter 1982). The first report of such patients was by Assal and colleagues (Assal *et al.* 1976) and has been followed by several others in the same decade (Assal *et al.* 1981; Landis *et al.* 1982; Van Lancker & Canter 1982; Van Lancker & Kreiman 1987; Van Lancker *et al.* 1988, 1989). Then, all interest in phonagnosia seems to have vanished (but see Peretz *et al.* 1994; Neuner & Schweinberger 2000) probably due to the lack of a standardized battery of voice discrimination and recognition, forcing interested researchers to devise their own tests.

Briefly, phonagnosia, like prosopagnosia the equivalent deficit for faces, has been found to occur most often after posterior right-hemisphere lesions. Phonagnosia has been dissociated from voice discrimination deficits (Van Lancker & Kreiman 1987) and doubly dissociated from aphasia; patients with receptive aphasia but unimpaired voice recognition have been reported, as well as patients with phonagnosia but normal speech perception (Van Lancker & Canter 1982). Most importantly for our discussion, at least one case of phonagnosia with preserved recognition of environmental sounds has been reported (Peretz *et al.* 1994), suggesting that voice recognition might rely on a different neural substrate than recognition of other sound sources, an argument for species specificity in the processing of voice. However, the poor resolution of the scanner used for lesion localization in most reported cases of phonagnosia prevents the precise neuroanatomical identification of these putative voice-specific mechanisms.

(f) *Neuroimaging evidence for voice-selective mechanisms in humans*

Neuroimaging studies using PET or functional magnetic resonance imaging (fMRI), by measuring non-invasively the cerebral activity of awake, behaving normal humans, have allowed substantial progress in our understanding of the functional organization of human auditory cortex (Zatorre & Binder 2000). In particular, a number of studies have investigated the neural correlates of speech perception and highlighted a large-scale network of parallel, distributed neuronal activity involving cortical regions, such as inferior prefrontal cortex, posterior temporal cortex, inferior parietal lobule and anterior STS with a predominance of the left hemisphere (Démonet *et al.* 1992; Binder *et al.* 2000; Scott *et al.* 2000; Crinion *et al.* 2003; Scott & Johnsrude 2003; but see Poeppel 1996; Price *et al.* 2005).

However, these studies typically contrasted speech stimuli with much lower-level control stimuli, such as tones, noise or amplitude-modulated noise. The lack of control stimuli of intermediary complexity makes it hard to understand exactly which features of the speech signal are responsible for the different components of

the pattern of cortical activation. Are these active regions really all involved in processing speech information? One troubling observation is that time-reversed speech—a signal which carries no linguistic content, although it essentially preserves the timbre and pitch variations of the voice—yields a pattern of activation quite similar to the one induced by the original speech signal in a large part of auditory cortex (Binder *et al.* 2000). Hence, could some lower-level features of the signal be determinant regardless of the speech content, for example, such as the signal's 'voiceness'?

(g) *Voice-selective areas along anterior STS*

Belin and colleagues used fMRI to compare the cortical activation patterns induced by vocal versus non-vocal sounds in normal adult volunteers (Belin *et al.* 2000, 2002; Fecteau *et al.* 2004, 2005). The vocal sounds were from a large variety of speakers spanning a large age range; they consisted of either speech sounds, such as syllables, words or connected speech in several languages, or non-speech vocal sounds, such as coughs, cries, laughs, various interjections, etc. The non-vocal sounds were matched in number, duration and energy, and consisted of instrumental, mechanical and environmental sounds or animal vocalizations. A first experiment used a block design and a passive listening task with only two categories: vocal and non-vocal.

In all participants, discrete regions of auditory cortex were found to respond significantly more to the vocal than to the non-vocal sounds (Belin *et al.* 2000). No region of auditory cortex was found to respond more to the non-vocal sounds. The anatomical localization of the voice-sensitive cortex was quite variable across subjects, unilateral on the left in some subjects, on the right in some others and bilateral in some (Belin *et al.* 2002), yet these regions were consistently located along the upper bank of the STS. The predominance of middle and anterior STS regions was confirmed in the group-level analysis. Interestingly, the voice-sensitive activity was the strongest on the right side, which appeared counter-intuitive at first, given the well-established advantage of the left hemisphere for speech (Belin *et al.* 2000; figure 2).

Follow-up experiments confirmed and extended the finding of voice-sensitivity along anterior STS. The voice-sensitive anterior STS regions were found to respond more strongly to voice than to control categories, such as a homogeneous category consisting of only bells or to acoustic control sounds equated in amplitude waveform or in an average long-term frequency (Belin *et al.* 2000). The STS voice-sensitive response therefore also proved to be quite selective. Fecteau *et al.* (2004) tested the species specificity of this response by comparing, using an event-related design, vocal and non-vocal sounds to a category of only cat vocalizations and a category of mixed animal vocalizations. The comparison of the human vocal with the non-vocal sounds again yielded bilateral activation along the middle and anterior STS; in contrast, the animal vocalizations, although matched in number and overall energy to the human vocalizations, only yielded marginal activation of the STS when compared with the non-vocal sounds (Fecteau *et al.* 2004).

(h) Electrophysiological evidence in humans

Electrophysiological techniques also proved useful in investigating the 'special' status of voice in the auditory cortical activity of normal, behaving adult humans. Levy *et al.* (2001) compared the evoked response potentials elicited by a sung voice and by pitch-matched notes played on different musical instruments. A late positive component peaking about 320 ms after sound onset was observed only in response to the sung voice (Levy *et al.* 2001). However, this 'voice-specific response' was not observed when participants did not attend to the auditory stimuli, or when they attended to features other than timbre. Thus, the 'voice-specific response' might reflect attentional processes related to the overriding salience of voice stimuli (Levy *et al.* 2003).

This important finding suggests an electrophysiological counterpart for the STS activations observed with fMRI. It is tempting to suggest that the generators of this late positivity may be located along anterior STS bilaterally. Yet 320 ms is a considerable time to show a differential response to such a biologically important sound category. It is a much longer time, for example, than the 170 ms that the visual cortex needs to differentiate faces from non-face objects, despite the later arrival of the sensory wave of information in visual compared with auditory cortex. Thus, one might reasonably make the hypothesis that components differentially sensitive to vocal and non-vocal information might be observed with earlier latencies, comparable to the face-selective N170. Such a putative early response remains to be discovered.

Overall, there is converging evidence from a variety of experimental techniques that a normal human brain contains several cortical areas selectively activated by sounds of human voice. This finding is very similar to the observations that several face-selective regions can be found in visual cortex (Puce *et al.* 1995; Kanwisher *et al.* 1997; Haxby *et al.* 2001), and suggests that face and voice processing could be organized following similar principles of cortical organizations (Belin *et al.* 2004). As for face processing, an important question arises: what is the functional role of the voice-selective cortical areas? Are they truly voice-selective? Or are they associated with our expertise for voices, and could be activated for other categories of expertise? This important question, still actively debated in the domain of face processing (Gauthier *et al.* 2000), is at present virtually unexplored in the domain of voice processing.

(i) Abnormal cortical response to voice in autism

Gervais *et al.* (2004) investigated the voice-sensitive cortical activity in autistic individuals. They used fMRI and the same protocol as Belin *et al.* (2000) to compare a group of five adults with autism with a group of eight age-matched controls. The control group showed an enhanced activation along anterior STS regions when vocal sounds were compared with non-vocal sounds, consistent with the previous experiments. In contrast, no voice-sensitive response could be observed in the autistic group (Gervais *et al.* 2004). When the responses to the vocal and non-vocal sounds were independently analysed, the response of the autistic group to the non-vocal sounds was found to be essentially normal, i.e. no different from that of the

control group. It is only for the vocal sounds that an abnormality appeared; the autistic participants failed to show additional STS activation for the sounds of voice. Their pattern of cerebral activation for the vocal sounds was essentially similar to that for the non-vocal sounds. In other words, for the auditory cortex of the autistic participants, voices had nothing special, they were just another sound category (Gervais *et al.* 2004).

The findings of Gervais *et al.* (2004) are interesting, in that the abnormal response of the cortex to sounds of voice is consistent with behaviour of autism and parallels recent findings of abnormal activation of face-processing networks in autism (Schultz *et al.* 2000). They also raise many questions that remain to be answered. One important question is whether the abnormal cortical response to voice can be generalized to all classes of vocal sounds and all groups of autistic subjects. Pelletier *et al.* (2005) recently investigated a small group of 'high functioning' autistic subjects, using the same experimental procedure as Gervais *et al.* (2004). This time, each autistic subject in whom functional images were successfully obtained showed activation of the STS in the vocal versus non-vocal comparison, comparable to the control subjects (Pelletier *et al.* 2005). Again, the small number of subjects calls for replication, but it seems that autism may not be automatically associated with abnormal cortical response to voice, and that variables such as performance IQ may prove to play a critical role. Future experiments need to investigate this possible relationship in more details and to relate cortical activity to measures of behavioural performance at voice perception tasks. In sum, the study of the neural correlates of voice perception in autism is a young but promising area of research which deserves as much attention as its counterpart in the domain of face processing.

4. PERCEPTION OF IDENTITY INFORMATION IN VOICE

It is a common observation that we can discriminate voices from different persons, extract much information on the physical characteristics of a speaker and often recognize familiar individuals from their voice alone. Do we share this ability to extract identity cues from voice with our non-human relatives? To begin with, are calls from different primates of a same species distinctive? If yes, do monkeys and apes actually use this identity information in their behaviour? And what are the neural correlates of these abilities?

(a) Identity information in primate vocalizations

The vocal production mechanism of primates allows a fair degree of variation in the acoustic structure of vocalizations, both inter-individually and across individuals. Slight differences in physical morphology between individuals of a same species have the potential to yield consistent acoustic differences. As discussed by Rendall *et al.* (1998), three main sources of individual variation that can lead to acoustic differences in a vocalization are as follows.

- (i) *Variation in laryngeal anatomy*, such as overall size of the larynx, size and relative proportion of

the vocal folds, amount of lubrication, presence of abnormalities, pattern of glottal closure, etc. Such variation in the source of the vocal tract can lead to differences in the fundamental frequency of phonation (f_0). Babies tend to have higher-pitched voice than adult females who tend to have higher-pitched voices than adult males (Titze 1989). Yet, the f_0 can also vary substantially in each individual, such that there is considerable overlap in f_0 range between these groups (Hillenbrand *et al.* 1995). Thus, f_0 alone is not a really good indicator of vocal identity (Kunzel 1989).

- (ii) *Variation in supralaryngeal anatomy*, such as in the shape and length of the vocal tract, elasticity of the tissues, etc. One particularly important parameter is the length of vocal tract, which is tightly related to the body size and largely determines the frequencies of the formants (Fitch 2000). However, the vocal tract length is not an absolute indicator of identity, since it can also show some degree of within-individual variability, particularly in humans; speech is essentially a rapid succession of fast changes of vocal tract shape that induce associated changes in formant frequencies. (Yet, individuals can be identified from sine-wave versions of their speech in which only formant frequencies are represented; Remez *et al.* 1997). Modifications of formant frequencies by alteration of the vocal tract length—such as by protruding lips—have also been observed in non-human primates, although the range of formant variation (the ‘vowel space’) is much smaller in non-human primates than in humans (Lieberman *et al.* 1969; Owren & Rendall 2003). The effect of the supralaryngeal filtering, and thus the perceptual salience of inter-individual variability, is the strongest for harmonically rich sounds, such as the coos or grunts of baboons (Owren *et al.* 1997) or the human vowels.
- (iii) *Variation in temporal patterning*, i.e. variations in the timing and duration of the vocalization that can be quite idiosyncratic and sometimes allow recognition, such as in some characteristic laughs that unmistakably identify their human owner.

Variations in voice production between individuals induce variations in the spectro-temporal distribution of acoustic energy, which in turn may or not have the potential to lead to successful discrimination or identification. In order to discriminate and eventually identify individuals based on their vocalizations, non-human primates as well as humans need to construct some sort of ‘vocal signature’, representation of an individual’s voice based on a combination of acoustic features that maximizes inter-individual variation while minimizing within-individual variation. It is clear to us that the human voice contains such combination of features, otherwise we would not be able to identify persons on the telephone. Is this the case as well for our non-human relatives? Can vocalizations from non-human primates be individually distinctive?

The information to discriminate speakers is indeed present in the vocalizations of many species of apes and monkeys as shown by several methods, in particular by statistical studies using clustering methods (reviewed in Snowdon 1986). Recent evidence for individual distinctiveness in vocalizations was obtained in the squirrel monkey (Boinski & Mitchell 1997), in the baboon (Owren *et al.* 1997), the rhesus monkey (Rendall *et al.* 1996, 1998; Owren & Rendall 2003), the Japanese macaque (Ceugniet & Izumi 2004a) and the cotton-top tamarin (Weiss *et al.* 2001). Owren *et al.* (1997) showed that the spectral energy peak (formant) patterning varied with caller identity in baboon grunts, and constituted the strongest grouping variable. The amplitude and frequency of the formants was found to emerge as a ‘predominant source of identity-based classificatory power’ (Owren *et al.* 1997). However, this information can be more present in some vocalizations than in others. Thus, the screams of rhesus monkeys appear to be less discriminative than the coos (Rendall *et al.* 1998), consistent with the idea that harmonically rich, low-frequency sounds comparable to our vowels are especially well suited to provide good estimates of vocal tract filtering effects. It has been suggested that these characteristics of the call may have been selected in the evolution partly owing to this reason (Brown 2003).

Thus, vocalizations by non-human primates can contain information that allows distinction of individuals. Can non-human primates use this information? Again, evidence using different methods in several species suggests that non-human primates, as humans, are able to use the idiosyncratic information in calls to discriminate or identify callers.

(b) Behavioural evidence in non-human primates

The ability to signal and perceive kin and identity at a distance through vocalizations—monkeys seem to avoid visual contacts in their social interactions—plays an important role in the social life of primates. It may constitute an adaptation of extreme importance in facilitating intra-group social cohesion (Rendall *et al.* 1996). Indeed, complex social interactions of most primates call for a good ability to discriminate between other group members from vocal cues alone, to extract kin relations, or even to explicitly recognize each other (Rendall *et al.* 1996, 1998).

Several studies investigated one particularly important example of vocal identification: the vocal recognition of infants by their mothers. The ability to accurately recognize her infant by his cries indeed provides a clear selective advantage by allowing the mother to respond appropriately to potentially dangerous situations, thus increasing offspring’s chances of survival. Kaplan *et al.* (1978) examined the responses of captive squirrel monkeys to vocalizations produced by their infants as well as by infants from other females. The responses of mothers to their own infant’s cries were clearly different from responses to cries from other infants, with a large increase in number of maternal vocalizations (Kaplan *et al.* 1978). Cheney & Seyfarth (1980), using playback of juvenile cries in a group of free-ranging vervet monkeys, found that the

mothers responded significantly faster and were more likely to approach the crying infant than other females (Cheney & Seyfarth 1980). Similar evidence was obtained in a group of three Japanese macaques mothers (Pereira 1986).

Other studies have investigated vocal recognition outside the mother–infant context. In an important study, Rendall *et al.* (1996) used single-trial playbacks in free-ranging rhesus macaques and showed that female rhesus responded faster and longer to contact ‘coo’ calls produced by a matrilineal relative than by a familiar, but non-kin individual (Rendall *et al.* 1996). Moreover, when tested with a habituation paradigm, the macaques showed a significant recovery from adaptation when the identity of the caller changed. Thus, these data suggest that monkey can extract enough information from a call to discriminate kin from non-kin individuals and discriminate between individuals. However, this ability does not generalize to all vocalizations, since screams were not found to allow accurate discrimination of kin or identity in a subsequent study (Rendall *et al.* 1998).

Comparable results were obtained in captive monkeys from several species. Weiss *et al.* (2001) used a habituation–dishabituation paradigm in the cotton-top tamarin and showed that habituation transferred when a different call was played but from the same individual, whereas they dishabituated when caller identity changed (Weiss *et al.* 2001). Ceugniet & Izumi (2004a,b) used an operant conditioning procedure to train two captive Japanese macaques to discriminate the calls from three conspecific callers (30 vocalizations each). The macaques then were able to successfully transfer discrimination of identity when new calls from these three callers were introduced (Ceugniet & Izumi 2004b). Interestingly, the monkeys performed less well, but still above chance, when the calls had been low-pass filtered to preserve only the first harmonic, thereby eliminating cues to vocal tract filtering. Thus, the patterning of source harmonics by the vocal tract is an important, but not essential cue to vocal identity; inter-individual variation related to the source or temporal patterning, which are comparatively preserved in the low-passed vocal stimuli, also allow above-than-chance identification of the ‘speaker’.

Evidence for identification of individual by vocal cues alone has also recently been obtained in greater apes. A captive female chimpanzee was shown to successfully match various calls (pan hoots, pan grunts and screams) from 10 different chimpanzee callers to the photograph of these callers (Kojima *et al.* 2003). She was also able to identify both callers of a duet of pan-hoots, suggesting that abilities of caller identification are present to a remarkable degree in chimpanzees. Thus, the available data show that non-human primates appear to be able to use the individually distinctive information present in voice to discriminate and recognize individuals.

(c) Behavioural evidence in humans

As we all can experience it each time we hear a voice, we are able to extract rich information on the physical characteristics and identity of a speaker/caller. An important corpus of studies has measured the accuracy

with which normal human listeners can extract different types of identity information (reviewed in Kreiman 1997).

The first physical characteristic we judge easily and relatively accurately is gender (Lass *et al.* 1976; Childers & Wu 1991; Wu & Childers 1991; Mullennix *et al.* 1995; Andrews & Schmidt 1997; Whiteside 1998a,b; Bachorowski & Owren 1999). Not very surprisingly, judgment of gender is quite accurate even in brief (Bachorowski & Owren 1999) or much degraded signals, such as in whispered speech (Tartter 1989) or sine-wave analogues of speech (Fellowes *et al.* 1997). The importance of not mistaking the gender of a potential mate clearly puts some evolutionary pressure to solve this particular ecological problem well. Childers & Wu (1991) and Bachorowski & Owren (1999) used statistical analyses of gender-related acoustical differences in voices, and showed that cues related both to the source (f_0) and to vocal tract characteristics (such as frequency of the second formant) were combined in an accurate representation of voice gender (Childers & Wu 1991; Bachorowski & Owren 1999).

Other physical characteristics can also be extracted with relatively good accuracy, although there is significant variation across listeners (Kreiman 1997). Estimated age is generally accurate within a decade, although listeners seem to underestimate the age of speakers (Hartman & Danaher 1976; Hartman 1979). Body size estimates have been found to be quite inaccurate, with a very small proportion of judgments actually correlating with the speakers’ height and weight (Lass & Davis 1976; Van Dommelen & Moxness 1995). Yet, listeners are found to be quite consistent in their judgement across several listening conditions, suggesting that vocal stereotypes are used to estimate body size, although these stereotypes are wrong (Gonzalez 2003). This inaccuracy is quite surprising as vocal tract length is well correlated with body size (at least in adult male humans; Rendall *et al.* 2005) and is tightly associated with formant frequencies (Fitch 1997), unlike the f_0 (Kunzel 1989; Rendall *et al.* 2005).

The ability to extract physical characteristics from voice peaks with identification of a speaker by the voice alone. A common finding is that some voices are easier to identify than others (Papcun *et al.* 1989; Kreiman 1997). Abberton & Fourcin (1978) reported above-than-chance accuracy in recognition of speakers from the output of a laryngograph, eliminating vocal tract contribution (cited in Kreiman 1997). Conversely, reliable speaker identification can be obtained from whispered speech (Tartter 1991), or sine-wave analogues of speech (Remez *et al.* 1997), demonstrating that, as for gender identification, acoustic cues related to both the laryngeal source and the supralaryngeal vocal tract are used to identify speakers. More research is now needed to understand how our perception of familiar and unfamiliar voices is organized in the brain, and which acoustic features are the most important in these processes.

As we have seen previously, the ability to extract information on the caller/speaker’s physical characteristics in voice is phylogenetically older than speech

perception, since we share it with other species, in particular other primates. There is also a strong evidence that voice perception abilities develop earlier than speech perception in ontogeny. Studies using measures of sucking preference or heart rate show that one-month-old infants prefer their mother's voice to voices from other persons (Mehler *et al.* 1978). This ability is even present in newborn babies (DeCasper & Fifer 1980) and extends to the father's voice (Ockleford *et al.* 1988). Recent measures in foetuses suggest that his ability is even present before birth (Kisilevsky *et al.* 2003). Thus, long before being able to discriminate and categorize the sounds of their maternal tongue, babies show impressive voice perception abilities.

In sum, humans clearly possess the ability to extract information on the physical characteristics and identity of a speaker from the voice alone. The voice recognition abilities of most normal listeners are clearly less accurate than face recognition, but still sufficient to extract useful information from an individual who is even out of sight. What are the neural correlates of the socially useful ability to recognize speakers?

(d) Neurobiological evidence in non-human primates

There are virtually no published data on the neural correlates of this ability to identify callers in non-human primates. Most researchers investigating neuronal responses to conspecific calls have focused on the issue of call selectivity (cf. §3*b*). The only evidence for neuronal mechanisms possibly involved in extracting vocal signatures of individual callers comes from outside of auditory cortex in the recent work by Romanski *et al.* (2004). These authors found that on the 301 auditory-responsive cells they recorded in macaque prefrontal cortex, two cells were found to be caller selective, i.e. to respond best to vocalizations from one caller and not to vocalizations from other callers (Romanski *et al.* 2004).

(e) Neurobiological evidence in humans

Despite the social importance of voice recognition abilities and the fact that they appeared earlier than speech in both phylogeny and ontogeny, little is known on their functional organization in the human brain. However, this ability has been the focus of several recent neuroimaging studies, and their results start shedding light on cortical regions involved in voice recognition abilities—which is comparatively much more than what is known in non-human primates (figure 3).

Imaimuzi and colleagues used PET to scan normal volunteers while listening to words pronounced by several actors and performing forced-choice identification of either the speaker pronouncing the words or the emotion that was portrayed in saying the word. The main finding was that the anterior temporal lobes were more active bilaterally during speaker identification than during emotion identification (Imaimuzi *et al.* 1997). A follow-up study compared a familiarity decision task on voices that were either familiar or unknown to the participants with a control phonetic decision task. Several cortical regions, including the

entorhinal cortex and the anterior part of the right temporal lobe, were found to be more active during the voice familiarity task. Moreover, cerebral blood flow in the right anterior temporal pole was correlated with the subjects' performance at a speaker identification task administered after scanning (Nakamura *et al.* 2001).

Belin & Zatorre (2003) used fMRI and a paradigm based on neuronal adaptation to investigate a putative representation of vocal signature in human auditory cortex. The reasoning was as follows: if a cortical region is involved in representing a speaker's vocal signature, then it should show adaptation or repetition-induced decrease in activity, in response to vocal samples produced by the same speaker—even if the samples correspond to different words. Normal volunteers were scanned while passively listening to auditory blocks corresponding to two conditions: in one condition (adapt-speaker), blocks were composed of 12 different syllables pronounced by a same speaker; in the control condition (adapt-syllable), blocks were composed of a same syllable spoken by 12 different speakers. The same 144 stimuli (12 syllables \times 12 speakers) were used in the two conditions, only the order of presentation changed. Owing to the similarities between the two conditions, most of the auditory cortices, including left and right A1, showed similar activation to the two conditions (Belin & Zatorre 2003). Only one region of auditory cortex showed a difference in activity between the two conditions; as predicted, this region showed significantly less activity in the adapt-speaker condition. Interestingly, it was located along the anterior STS in the right hemisphere, just a few millimetres away from one of the maxima of voice-selectivity previously observed (Belin *et al.* 2000).

A remarkably convergent finding was obtained by another team using a nearly opposite design: whereas Belin & Zatorre (2003) used a bottom-up design, manipulating stimuli but not task, von Kriegstein *et al.* (2003) scanned normal volunteers while they were attending either to the linguistic content of German utterances or to the speaker of these same utterances (von Kriegstein *et al.* 2003). They found that the right anterior STS and a part of the right precuneus were more active when the identification task was focused on the speaker's identity, whereas a left middle STS region was more active in the reverse comparison. Thus, although the vocal stimuli were similar in the two conditions, directing attention to vocal identity was found to increase activity in a region of right anterior STS very close to that observed by Belin & Zatorre (2003). Using complementary analyses, von Kriegstein & Giraud (2004) further documented the functional organization of right STS. When comparing the responses to familiar versus unfamiliar voices, they outlined a region of the posterior part of the STS that responded more during speaker recognition when the voices were unfamiliar (von Kriegstein & Giraud 2004). Functional connectivity analyses showed that both anterior and posterior regions of the right STS interacted with a more central part of right STS located close to the maxima of sensitivity to the acoustic structure of voice (von Kriegstein & Giraud 2004).

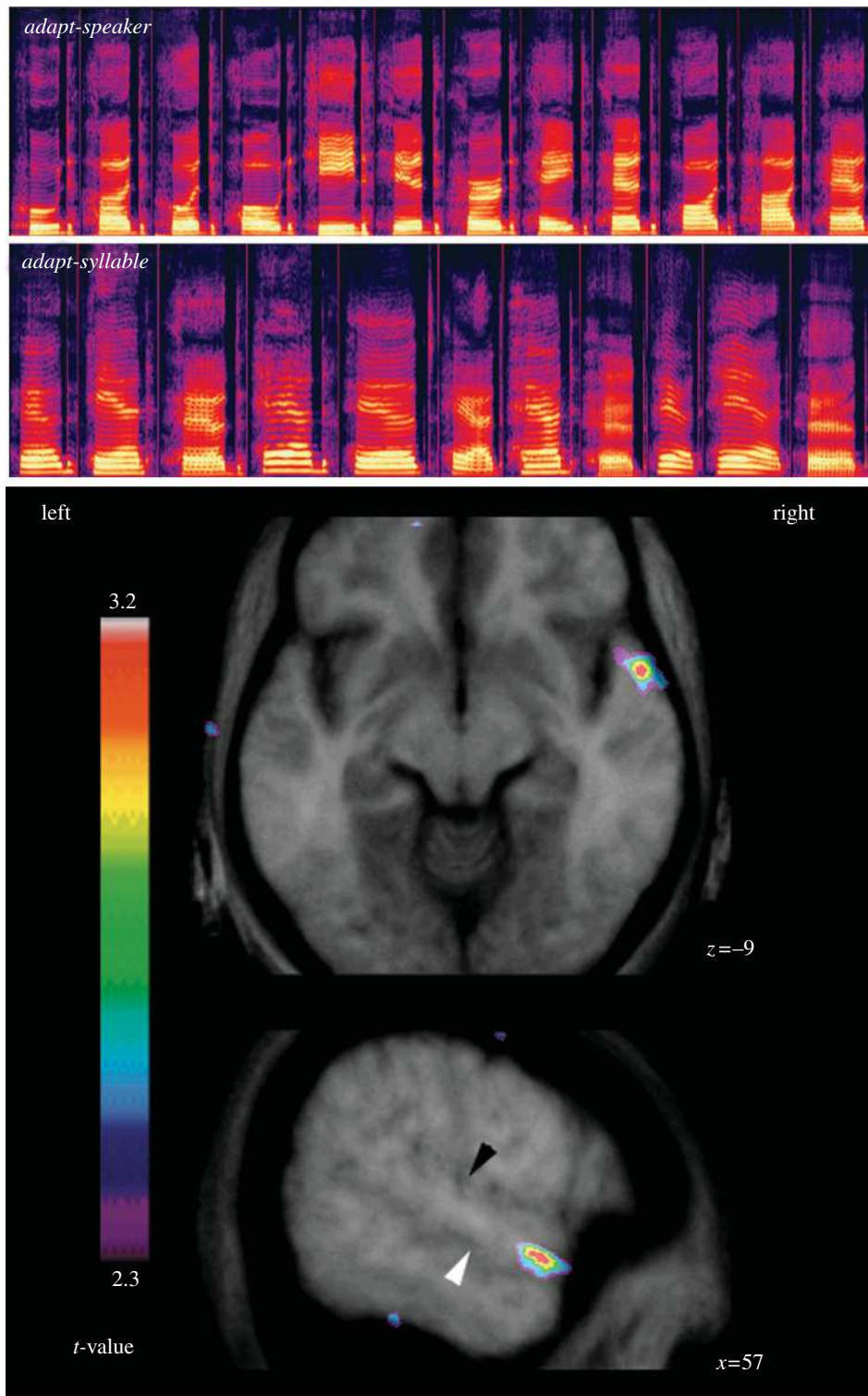


Figure 3. Cortical sensitivity to vocal identity. (a) Spectrograms (0–5000 Hz) of examples of auditory blocks used by Belin & Zatorre (2003). *Adapt-speaker*: different syllables spoken by a same speaker. *Adapt-syllable*: a same syllable spoken by several different speakers. (b) Cortical regions showing decrease in neuronal activity with repetition of the speaker's voice, shown in colour scale on axial (top) and sagittal (middle) slices through the subjects' mean anatomical image. Reproduced with permission from Belin & Zatorre (2003).

5. CONCLUSIONS

Converging evidence from human studies clearly points to an important role of anterior STS regions in processing voice information, particularly related to speaker's identity, with a clear functional lateralization to the right hemisphere. These findings in humans allow two important conclusions for studies in non-human primates.

First, single-cell recordings focusing on anterior STS regions would probably yield highly interesting findings. The STS regions play a clear role in human voice perception in particular and social cognition in general (Allison *et al.* 2000). Evidence in the macaque shows that these regions also contain neurons selectively tuned to faces (Perrett *et al.* 1992) and to sounds of actions (Barraclough *et al.* 2005). Recent data

suggest that some of these STS regions (particularly, area TAA) may send direct projections to the auditory prefrontal domain identified in the macaque brain (L. M. Romanski 2006, personal communication), further suggesting that STS regions would be particularly well suited to combine information from vocalizations and face displays, and yield a supra-modal representation of conspecific individuals.

A second important conclusion relates to the patterns of functional lateralization. The well-known advantage of the left hemisphere in processing speech information has pushed researchers to look for a comparable left-hemisphere advantage for processing communication sounds in non-human animals. Yet what the human studies suggest is that the left-hemisphere advantage only holds when processing speech information. Human studies having manipulated subject's attention towards non-linguistic features of the vocal signal, such as prosody (Zatorre *et al.* 1992) or speaker identity (von Kriegstein *et al.* 2003), clearly showed that a right-hemisphere advantage can be obtained. Thus, patterns of functional lateralization in non-human primates may also not be exclusively biased towards the left hemisphere, particularly when attending to caller affect or identity.

REFERENCES

- Abberton, E. & Fourcin, A. 1978 Intonation and speaker identification. *Lang. Speech* **21**, 305–318.
- Aitkin, L. M., Merzenich, M. M., Irvine, D. R., Clarey, J. C. & Nelson, J. E. 1986 Frequency representation in auditory cortex of the common marmoset (*Callithrix jacchus jacchus*). *J. Comp. Neurol.* **252**, 175–185. (doi:10.1002/cne.902520204)
- Allison, T., Puce, A. & McCarthy, G. 2000 Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4**, 267–278. (doi:10.1016/S1364-6613(00)01501-1)
- Andrews, M. L. & Schmidt, C. P. 1997 Gender presentation: perceptual and acoustical analyses of voice. *J. Voice* **11**, 307–313. (doi:10.1016/S0892-1997(97)80009-4)
- Assal, G., Zander, E., Kremin, H. & Buttet, J. 1976 Discrimination des voix lors des lésions du cortex cérébral. *Arch. Suisses Neurol. Neurochir. Psychiatrie* **119**, 307–315.
- Assal, G., Aubert, C. & Buttet, J. 1981 Asymétrie cérébrale et reconnaissance de la voix. *Revue Neurolog.* **137**, 255–268.
- Bachorowski, J. A. & Owren, M. J. 1999 Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J. Acoust. Soc. Am.* **106**, 1054–1063. (doi:10.1121/1.427115)
- Barralough, N. E., Xiao, D., Baker, C. I., Oram, M. W. & Perrett, D. I. 2005 Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J. Cogn. Neurosci.* **17**, 377–391. (doi:10.1162/0898929053279586)
- Belin, P. & Zatorre, R. J. 2003 Adaptation to speaker's voice in right anterior temporal-lobe. *Neuroreport* **14**, 2105–2109. (doi:10.1097/00001756-200311140-00019)
- 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. *Cogn. Brain Res.* **13**, 17–26. (doi:10.1016/S0926-6410(01)00084-2)
- Belin, P., Fecteau, S. & Bédard, C. 2004 Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* **8**, 129–135. (doi:10.1016/j.tics.2004.01.008)
- Binder, J. R. *et al.* 2000 Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528.
- Boinski, S. & Mitchell, C. L. 1997 Chuck vocalizations of wild female squirrel monkeys (*Saimiri sciureus*) contain information on caller identity and foraging activity. *Int. J. Primatol.* **18**, 975–993. (doi:10.1023/A:1026300314739)
- Brown, C. H. 2003 Ecological and physiological constraints for primate vocal communication. In *Primate audition. Ethology and neurobiology* (ed. A. A. Ghazanfar), pp. 127–149. Boca Raton, FL: CRC Press.
- Ceugniet, M. & Izumi, A. 2004a Individual vocal differences of the coo call in Japanese monkeys. *CR Biol.* **327**, 149–157.
- Ceugniet, M. & Izumi, A. 2004b Vocal individual discrimination in Japanese monkeys. *Primates* **45**, 119–128.
- Cheney, D. L. & Seyfarth, R. M. 1980 Vocal recognition in free-ranging vervet monkeys. *Anim. Behav.* **28**, 362–367. (doi:10.1016/S0003-3472(80)80044-3)
- Childers, D. G. & Wu, K. 1991 Gender recognition from speech. Part II: fine analysis. *J. Acoust. Soc. Am.* **90**, 1841–1856. (doi:10.1121/1.401664)
- Coslett, H. B., Brashear, H. R. & Heilman, K. M. 1984 Pure word deafness after bilateral primary auditory cortex infarcts. *Neurology* **34**, 347–352.
- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D. & Wise, R. J. 2003 Temporal lobe regions engaged during normal speech comprehension. *Brain* **126**, 1193–1201. (doi:10.1093/brain/awg104)
- Damasio, A. R. 1992 Aphasia. *N. Engl. J. Med.* **326**, 531–539.
- DeCasper, A. J. & Fifer, W. P. 1980 Of human bonding: newborns prefer their mothers' voices. *Science* **208**, 1174–1176. (doi:10.1126/science.7375928)
- Démonet, J. F. *et al.* 1992 The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768.
- Eskenazi, L., Childers, D. G. & Hicks, D. M. 1990 Acoustic correlates of vocal quality. *J. Speech Hear Res.* **33**, 298–306.
- Fant, G. 1960 *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
- Farah, M. J. 1996 Is face recognition 'special'? Evidence from neuropsychology. *Behav. Brain Res.* **76**, 181–189.
- Fecteau, S., Armony, J., Joannette, Y. & Belin, P. 2004 Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* **23**, 840–848. (doi:10.1016/j.neuroimage.2004.09.019)
- Fecteau, S., Armony, J., Joannette, Y. & Belin, P. 2005 Sensitivity to voice in human prefrontal cortex. *J. Neurophysiol.* **94**, 2251–2254. (doi:10.1152/jn.00329.2005)
- Fellowes, J. M., Remez, R. E. & Rubin, P. E. 1997 Perceiving the sex and identity of a talker without natural vocal timbre. *Percept. Psychophys.* **59**, 839–849.
- 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. (doi:10.1121/1.421048)
- Fitch, W. T. 2000 The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267. (doi:10.1016/S1364-6613(00)01494-7)
- Fitch, W. T. 2003 Primate vocal production and its implications for auditory research. In *Primate audition. Ethology and neurobiology* (ed. A. A. Ghazanfar), pp. 1–12. Boca Raton, FL: CRC Press.
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C. & Anderson, A. W. 2000 The fusiform "face area" is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* **12**, 495–504. (doi:10.1162/089892900562165)

- Gervais, H. *et al.* 2004 Abnormal voice processing in autism: a fMRI study. *Nat. Neurosci.* **7**, 801–802. (doi:10.1038/nl1291)
- Ghazanfar, A. A. & Santos, L. R. 2003 Primates as auditory specialists. In *Primate audition. Ethology and neurobiology* (ed. A. A. Ghazanfar), pp. 1–12. Boca Raton, FL: CRC Press.
- 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. (doi:10.1159/000047270)
- 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. *Proc. Natl Acad. Sci. USA* **101**, 17 516–17 521. (doi:10.1073/pnas.0408077101)
- 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. (doi:10.1016/0378-5955(83)90131-4)
- Gonzalez, J. 2003 Estimation of speakers' weight and height from speech: a re-analysis of data from multiple studies by Lass and colleagues. *Percept. Mot. Skills* **96**, 297–304. (doi:10.2466/PMS.96.1.297-304)
- Green, S. 1975 Variation of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): a field study. In *Primate behavior*, vol. 4 (ed. E. L. A. Rosenblum), pp. 1–102. New York, NY: Academic Press.
- Hartman, D. E. 1979 The perceptual identity and characteristics of aging in normal male adult speakers. *J. Commun. Disord.* **12**, 53–61. (doi:10.1016/0021-9924(79)90021-2)
- Hartman, D. E. & Danahuer, J. L. 1976 Perceptual features of speech for males in four perceived age decades. *J. Acoust. Soc. Am.* **59**, 713–715.
- Hauser, M. 1996 *The evolution of communication*. Cambridge, MA: MIT Press.
- 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. USA* **91**, 3946–3948. (doi:10.1073/pnas.91.9.3946)
- Hauser, M. D., Agnetta, B. & Perez, C. 1998 Orienting asymmetries in rhesus monkeys: the effect of time-domain changes on acoustic perception. *Anim. Behav.* **56**, 41–47. (doi:10.1006/anbe.1998.0738)
- Hauser, M. D., Chomsky, N. & Fitch, W. T. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579.
- 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. (doi:10.1126/science.1063736)
- Heffner, H. E. & Heffner, R. S. 1984 Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* **226**, 75–76. (doi:10.1126/science.6474192)
- Heffner, H. E. & Heffner, R. S. 1986 Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *J. Neurophysiol.* **56**, 683–701.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J. & Wheeler, K. 1995 Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**, 1300–1313.
- Hopp, S. L., Sinnott, J. M., Owren, M. J. & Petersen, M. R. 1992 Differential sensitivity of Japanese macaques (*Macaca fuscata*) and humans (*Homo sapiens*) to peak position along a synthetic coo call continuum. *J. Comp. Psychol.* **106**, 128–136. (doi:10.1037/0735-7036.106.2.128)
- Imaizumi, S. *et al.* 1997 Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* **8**, 2809–2812.
- Kaas, J. H. & Hackett, T. A. 1999 'What' and 'where' processing in auditory cortex. *Nat. Neurosci.* **2**, 1045–1047. (doi:10.1038/15967)
- Kanwisher, N., McDermott, J. & Chun, M. M. 1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311.
- Kaplan, J. N., Winship-Ball, A. & Sim, L. 1978 Maternal discrimination of infant vocalizations in squirrel monkeys. *Primates* **19**, 187–193.
- Kisilevsky, B. S. *et al.* 2003 Effects of experience on fetal voice recognition. *Psychol. Sci.* **14**, 220–224. (doi:10.1111/1467-9280.02435)
- Klatt, D. H. 1980 Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* **67**, 971–995. (doi:10.1121/1.383940)
- Kojima, S., Izumi, A. & Ceugniet, M. 2003 Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates* **44**, 225–230.
- Kreiman, J. 1997 Listening to voices: theory and practice in voice perception research. In *Talker variability in speech research* (ed. K. Johnson & J. Mullenix), pp. 85–108. New York, NY: Academic Press.
- Kunzel, H. J. 1989 How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* **46**, 117–125.
- Landis, T., Buttet, J., Assal, G. & Graves, R. 1982 Dissociation of ear preference in monaural word and voice recognition. *Neuropsychologia* **20**, 501–504. (doi:10.1016/0028-3932(82)90049-5)
- Lass, N. J. & Davis, M. 1976 An investigation of speaker height and weight identification. *J. Acoust. Soc. Am.* **60**, 700–703.
- Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T. & Bourne, V. T. 1976 Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* **59**, 675–678. (doi:10.1121/1.380917)
- Levy, D. A., Granot, R. & Bentin, S. 2001 Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport* **12**, 2653–2657. (doi:10.1097/00001756-200108280-00013)
- Levy, D. A., Granot, R. & Bentin, S. 2003 Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology* **40**, 291–305.
- 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. (doi:10.1126/science.164.3884.1185)
- Manley, J. A. & Muller-Preuss, P. 1978 Response variability of auditory cortex cells in the squirrel monkey to constant acoustic stimuli. *Exp. Brain Res.* **32**, 171–180. (doi:10.1007/BF00239725)
- Mehler, J., Bertoncini, J. & Barriere, M. 1978 Infant recognition of mother's voice. *Perception* **7**, 491–497.
- Mullenix, 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. (doi:10.1121/1.413832)
- Nakamura, K. *et al.* 2001 Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* **39**, 1047–1054. (doi:10.1016/S0028-3932(01)00037-9)
- Neuner, F. & Schweinberger, S. R. 2000 Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn.* **44**, 342–366.

- Newman, J. D. 2003 Auditory communication and mechanisms in the squirrel monkey. In *Primate audition. Ethology and neurobiology* (ed. A. A. Ghazanfar), pp. 27–246. Boca Raton, FL: CRC Press.
- Ockleford, E. M., Vince, M. A., Layton, C. & Reader, M. R. 1988 Responses of neonates to parents' and others' voices. *Early Hum. Dev.* **18**, 27–36. (doi:10.1016/0378-3782(88)90040-0)
- Owren, M. J. & Linker, C. D. 1995 Some analysis methods that may be useful to acoustic primatologists. In *Current topics in primate vocal communication* (ed. E. Zimmermann, J. D. Newman & U. Jürgens), pp. 1–27. New York, NY: Plenum.
- Owren, M. J. & Rendall, D. 2003 Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: perceptual experiments with human (*Homo sapiens*) listeners. *J. Comp. Psychol.* **117**, 380–390. (doi:10.1037/0735-7036.117.4.380)
- Owren, M. J., Hopp, S. L., Sinnott, J. M. & Petersen, M. R. 1988 Absolute auditory thresholds in three Old World monkey species (*Cercopithecus aethiops*, *C. neglectus*, *Macaca fuscata*) and humans (*Homo sapiens*). *J. Comp. Psychol.* **102**, 99–107. (doi:10.1037/0735-7036.102.2.99)
- Owren, M. J., Seyfarth, R. M. & Hopp, S. L. 1992 Categorical vocal signaling in nonhuman primates. In *Nonverbal vocal communication. Comparative and developmental approaches* (ed. H. Papoušek, U. Jürgens & M. Papoušek), pp. 102–122. Cambridge, UK: Cambridge University Press. Editions de la Maison des Sciences de l'Homme
- Owren, M. J., Seyfarth, R. M. & Cheney, D. L. 1997 The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions. *J. Acoust. Soc. Am.* **101**, 2951–2963. (doi:10.1121/1.418523)
- Papcun, G., Kreiman, J. & Davis, A. 1989 Long-term memory for unfamiliar voices. *J. Acoust. Soc. Am.* **85**, 913–925. (doi:10.1121/1.397564)
- Pelletier, I., Belin, P., Gervais, H., Jemel, B., Mottron, L. & Zilbovicius, M. 2005 *Typical brain activation for voices in high functioning autistic individuals; a preliminary study*. Paper presented at the International Meeting for Autism Research, Boston, MA.
- Pereira, M. E. 1986 Maternal recognition of juvenile offspring coo vocalizations in Japanese macaques. *Anim. Behav.* **34**, 935–937. (doi:10.1016/S0003-3472(86)80084-7)
- Peretz, I. *et al.* 1994 Functional dissociations following bilateral lesions of auditory cortex. *Brain* **117**, 1283–1301.
- Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. 1992 Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. R. Soc. B* **335**, 23–30.
- 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**(4365), 324–327. (doi:10.1126/science.99817)
- Petersen, M. R. *et al.* 1984 Neural lateralization of vocalizations by Japanese macaques: communicative significance is more important than acoustic structure. *Behav. Neurosci.* **98**, 779–790. (doi:10.1037/0735-7044.98.5.779)
- Poeppl, D. 1996 A critical review of PET studies of phonological processing. *Brain Lang.* **55**, 317–351. (doi:10.1006/brln.1996.0108)
- 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. (doi:10.1038/nature02268)
- Price, C., Thierry, G. & Griffiths, T. D. 2005 Speech-specific auditory processing: where is it? *Trends Cogn. Sci.* **9**, 271–276. (doi:10.1016/j.tics.2005.03.009)
- Puce, A., Allison, T., Gore, J. C. & McCarthy, G. 1995 Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* **74**, 1192–1199.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D. & Mehler, J. 2000 Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* **288**, 349–351. (doi:10.1126/science.288.5464.349)
- Rauschecker, J. P. & Tian, B. 2000 Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11 800–11 806. (doi:10.1073/pnas.97.22.11800)
- Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrel, T. D. 1981 Speech perception without traditional speech cues. *Science* **212**, 947–950. (doi:10.1126/science.7233191)
- Remez, R. E., Fellowes, J. M. & Rubin, P. E. 1997 Talker identification based on phonetic information. *Exp. Psychol. Hum. Percept. Perform.* **23**, 651–666. (doi:10.1037/0096-1523.23.3.651)
- 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. (doi:10.1006/anbe.1996.0103)
- Rendall, D., Owren, M. J. & Rodman, P. S. 1998 The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* **103**, 602–614. (doi:10.1121/1.421104)
- Rendall, D., Kollias, S., Ney, C. & Lloyd, P. 2005 Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *J. Acoust. Soc. Am.* **117**, 944–955. (doi:10.1121/1.1848011)
- Romanski, L. M. & Goldman-Rakic, P. S. 2002 An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* **5**, 15–16. (doi:10.1038/nn781)
- Romanski, L. M., Averbeck, B. B. & Diltz, M. 2004 Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* **93**, 734–747. (doi:10.1152/jn.00675.2004)
- Schön Ybarra, M. A. 1995 A comparative approach to the non-human primate vocal tract: implications for sound production. In *Current topics in primate vocal communication* (ed. E. Zimmermann, J. D. Newman & U. Jürgens), pp. 185–198. New York, NY: Plenum Press.
- Schultz, R. T. *et al.* 2000 Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. *Arch. Gen. Psychiatry* **57**, 331–340. (doi:10.1001/archpsyc.57.4.331)
- Scott, S. K. & Johnsrude, I. S. 2003 The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107. (doi:10.1016/S0166-2236(02)00037-1)
- Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. 2000 Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406. (doi:10.1093/brain/123.12.2400)
- Seyfarth, R. M., Cheney, D. L. & Marler, P. 1980 Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* **210**, 801–803. (doi:10.1126/science.7433999)
- Shoumaker, R. D., Ajax, E. T. & Schenkenberg, T. 1977 Pure word deafness. (Auditory verbal agnosia). *Dis. Nerv. Syst.* **38**, 293–299.
- Sinnott, J. M. 1989 Detection and discrimination of synthetic english vowels by old world monkeys (*Cercopithecus*, *Macaca*) and humans. *J. Acoust. Soc. Am.* **86**, 557–565. (doi:10.1121/1.398235)

- Sinnott, J. M. & Adams, F. S. 1987 Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *J. Acoust. Soc. Am.* **82**, 1539–1547. (doi:10.1121/1.395144)
- Snowdon, C. T. 1986 Vocal communication. In *Comparative primate biology: behaviour, conservation and ecology* (ed. G. Mitchell & J. Erwin), pp. 495–530. New York, NY: Alan R. Liss.
- Tartter, V. C. 1989 What's in a whisper? *J. Acoust. Soc. Am.* **86**, 1678–1683. (doi:10.1121/1.398598)
- Tartter, V. C. 1991 Identifiability of vowels and speakers from whispered syllables. *Percept. Psychophys.* **49**, 365–372.
- Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. 2001 Functional specialization in rhesus monkey auditory cortex. *Science* **292**, 290–293. (doi:10.1126/science.1058911)
- Titze, I. R. 1989 Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* **85**, 1699–1707. (doi:10.1121/1.397959)
- Titze, I. R. 1993 Current topics in voice production mechanisms. *Acta Otolaryngol.* **113**, 421–427.
- Ungerleider, L. G. & Haxby, J. V. 1994 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* **4**, 157–165. (doi:10.1016/0959-4388(94)90066-3)
- Van Dommelen, W. A. & Moxness, B. H. 1995 Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Lang. Speech* **38**, 267–287.
- Van Lancker, D. R. & Canter, G. J. 1982 Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn.* **1**, 185–195. (doi:10.1016/0278-2626(82)90016-1)
- Van Lancker, D. & Kreiman, J. 1987 Voice discrimination and recognition are separate abilities. *Neuropsychologia* **25**, 829–834. (doi:10.1016/0028-3932(87)90120-5)
- 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. *J. Clin. Exp. Neuropsychol.* **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. (doi:10.1016/j.neuroimage.2004.02.020)
- von Kriegstein, K., Eger, E., Kleinschmidt, A. & Giraud, A. L. 2003 Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* **17**, 48–55. (doi:10.1016/S0926-6410(03)00079-X)
- Wang, X. 2000 On cortical coding of vocal communication sounds in primates. *Proc. Natl Acad. Sci. USA* **97**, 11 843–11 849. (doi:10.1073/pnas.97.22.11843)
- 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. Neurophysiol.* **74**, 2685–2706.
- Weiss, D. J., Garibaldi, B. T. & Hauser, M. D. 2001 The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): acoustic analyses and playback experiments. *J. Comp. Psychol.* **115**, 258–271. (doi:10.1037/0735-7036.115.3.258)
- Wernicke, C. 1874 *Der aphasische Symptomencomplex*. Breslau, Germany: Cohn & Weigert.
- Whiteside, S. P. 1998a Identification of a speaker's sex: a fricative study. *Percept. Mot. Skills* **86**, 587–591.
- Whiteside, S. P. 1998b Identification of a speaker's sex: a study of vowels. *Percept. Mot. Skills* **86**, 579–584.
- 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. (doi:10.1007/BF00234133)
- Winter, P., Ploog, D. & Latta, J. 1966 Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Exp. Brain Res.* **1**, 359–384. (doi:10.1007/BF00237707)
- Wollberg, Z. & Newman, J. D. 1972 Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. *Science* **175**, 212–214.
- Wu, K. & Childers, D. G. 1991 Gender recognition from speech. Part I: coarse analysis. *J. Acoust. Soc. Am.* **90**, 1828–1840. (doi:10.1121/1.401663)
- Zatorre, R. J. & Binder, J. R. 2000 Functional and structural imaging of the human auditory system. In *Brain mapping: the systems* (ed. A. Toga & J. Mazziota), pp. 365–402. Los Angeles, CA: Academic Press.
- Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. 1992 Lateralisation of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849. (doi:10.1126/science.1589767)
- Zoloth, S. R. *et al.* 1979 Species-specific perceptual processing of vocal sounds by monkeys. *Science* **204**, 870–873. (doi:10.1126/science.108805)