Review TRENDS in Cognitive Sciences Vol.8 No.3 March 2004

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Thinking the voice: neural correlates of voice perception

Pascal Belin, Shirley Fecteau and Catherine Bédard

Laboratoire de neuro-cognition vocale, Groupe de recherche en neuropsychologie et cognition (GRENEC), Département de Psychologie, Université de Montréal, CP 6128 succ. centre-ville, Montréal H3C 3J7, Québec, Canada

The human voice is the carrier of speech, but also an 'auditory face' that conveys important affective and identity information. Little is known about the neural bases of our abilities to perceive such paralinguistic information in voice. Results from recent neuroimaging studies suggest that the different types of vocal information could be processed in partially dissociated functional pathways, and support a neurocognitive model of voice perception largely similar to that proposed for face perception.

The human voice is the most important sound of our auditory environment. We probably spend more time everyday listening to voices than to any other sound, and our ability to analyze and categorize information contained in voices plays a key role in human social interactions. Voice is of course the carrier of speech, but there is more to voice than 'simply' speech. Speech appeared recently in evolution as a particularly complex and abstract use of voice by the human species [1,2]. However, vocalizations were prominent in the auditory environment of vertebrates for millions of years before speech emerged. Accurately perceiving the information contained in vocalizations from conspecific individuals, prey or predators is of crucial importance for survival. Like many other species, we are endowed with abilities to extract 'paralinguistic' information in voices (see Box 1). For example, even when speech information is not available in a voice – because it is a baby cry, or a cough, or heard through a wall or at a distance – we are still able to extract valuable information about the identity and the affective state of the person who produces the vocalization.

The abilities involved in perceiving paralinguistic information in voices – or 'voice perception' abilities – have been far less investigated than speech perception, and little is known about their neural bases. Results from recent neuroimaging studies, however, suggest that the different types of vocal information could be processed in partially dissociated functional pathways. Because speech emerged when cerebral mechanisms already existed for analyzing other types of vocal information, studying speech perception in the broader context of voice perception might provide a useful perspective.

The voice: an 'auditory face'

The voice not only contains *speech information*, it can also be viewed as an 'auditory face', that allows us to recognize individuals and emotional states. Voices, as faces, are characterized by a unique combination of physical features related to the unique configuration of the human vocal apparatus (see Box 2). As with faces, minute inter- and intra-individual variations around that generic structure carry useful information, which can be divided in three broad categories: speech information, but also identity information and affective information [3].

The voice carries important *identity information* in 'invariant' or 'static' features such as timbre – directly influenced by physical factors such as age and gender – and also in 'dynamic' information, such as patterns of pronunciation specific to a region (accent) or to a person – there are such unique laughs… Listeners are generally accurate at determining the gender of the speaker [4,5] and his/her approximate age [6,7]. The ability to judge other physical characteristics such as height, weight, racial group or even psychological characteristics, such as trustworthiness, is more controversial [8]. Our ability to use identity information culminates with speaker recognition: even after long periods of time, we can recognize persons from their voice with surprising accuracy [9,10].

The voice also contains *affective information:* as with faces, voices are directly influenced by the speaker's affective state. The modification of acoustic parameters induced by autonomic influence and specific patterns of muscular contraction corresponding to various affective

Box 1. Ontogenesis and phylogenesis of voice perception

Human babies cannot talk or understand speech, yet they are able to recognize voice. Experiments measuring changes in heart rates in neonates during presentation of different voices demonstrate an ability to discriminate voices, and to recognize the voices of their mother and father [72,73]. This ability is apparently even present in term fetuses before birth [74,75].

Speech perception might be unique to humans, but other voice perception abilities exist in other species. For example, the ability to identify conspecific individuals based on their vocalizations has been demonstrated in macaques [76,77]. A particularly striking example of vocal recognition is the case of the northern fur seal: not only do pups and mothers have the ability to recognize each other's vocalizations during the breeding season, despite the large population of the colony, but they are also able to retain these memories for at least 4 years [78]; this learning appears to occur as rapidly as the first 2–5 days of life [79].

Corresponding author: Pascal Belin (pascal.belin@umontreal.ca).

 1364-6613/$ - see front matter © 2004 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2004.01.008

Box 2. Variability in voice quality

The voice is the result of a source in the larynx filtered by the supra-laryngeal vocal tract ('source/filter theory' [2,80]). The periodic opening and closing of the vocal folds in the larynx produces a buzzing sound with a characteristic waveform, that determines the fundamental frequency (f0) of phonation – as well as a variable amount of aspiration noise that contributes to the 'breathy' and 'whispery' qualities of voice. Besides the modal register used in normal speaking or singing, the larynx can also be used in the 'fry mode' – with lowest f0 and sounding like growling or groaning – and in the 'falsetto mode' – which sounds thinner and can produce the highest notes. The vocal tract can be viewed as a complex mobile filter that enhances certain frequencies of the laryngeal source and attenuates others. Enhanced frequencies, called formants, depend both on the size of the vocal tract – correlated to individual body size, unlike f0 [15] – and on its shape, determined by the configuration of the articulators.

Gender and age differences in size of the larynx and vocal tract strongly influence voice quality: men have lower average f0 and formant frequencies – although range of f0 largely overlap for men and women. Women use their vocal folds in a more open configuration than men, leading to a higher ratio of low to high frequencies and more aspiration noise [81–83]. Cultural factors and vocal habits also play a major role in shaping voice quality. For example, gender can be well identified from voice alone in pre-pubertal children even though there are no known sex differences in the anatomy of the larynx and vocal tract at this age [84,85].

states allow us to gather important information on a person's emotional and motivational state. Perception of this information in voice has mostly been studied in the context of speech [11]. Emotional prosody – a set of acoustic parameters of speech directly influenced by affect such as mean amplitude, segment and pause duration, mean fundamental frequency (f0) and f0 variation – allows the listener to infer much of the speaker's affective state [12]. Non-speech interjections – such as laughs, cries, screams and moans – also contain rich affective information, and can be viewed as the auditory equivalent of facial emotional expressions (that most often accompany them). As in the case of identity information, affective information can be contained both in 'static' features – such as the characteristic timbre of a screaming voice – or in more 'dynamic' features, such as the melodic contour of an utterance.

Importantly, vocal features can carry more than one type of information. Speech formants are a good example because they carry both speech and identity information. Formants correspond to the frequencies that are amplified by supra-laryngeal filtering and convey important phonetic information: most voiced phonemes can be approximated well by synthesizing sounds with energy at formant frequencies [13]. Although they lack vocal quality, 'caricatures' of speech sentences composed of pure tones at the first three formant frequencies ('sine-wave speech') can be understood well [14]. Formants also carry important identity information: they are directly related to the size of the vocal tract and can therefore provide estimates of body size [15], and familiar speakers can be identified from sine-wave analogues of their vocalizations [16].

In the domain of face processing, functional models have suggested a partial dissociation of the neural processing of these three types of information. It is tempting to suggest that the neural substrate for processing vocal information could be organized following similar principles, and that speech, affective and identity information in voice could be processed in partially dissociated functional pathways (see Box 3). In the next section, we review recent neuroimaging studies of paralinguistic processing of voice, with emphasis on the less-explored field of speaker recognition.

Neural correlates of voice perception

Perception of speech information

Most neuroimaging studies [21–24] in the voice domain have investigated some aspects of the functional architecture involved in speech perception. These studies, reviewed elsewhere [17–20], have outlined the specific involvement in speech perception of bilateral, non-primary regions of the superior temporal cortex, both posterior (planum temporale) and anterior to Heschl's gyrus, extending inferiorly to the middle and anterior parts of superior temporal sulcus (STS). Several studies have suggested a dissociation between middle STS regions, more responsive to the presence of speech but not to the meaning (e.g. response to backwards speech but not to understandable modulated noise), and more anterior regions of the left STS/superior temporal plane, which seem to be more involved in comprehension, even from an input with a much degraded acoustic structure.

Perception of vocal affective information

Fewer studies have used neuroimaging techniques to investigate the perception of affective information in voice. Most of these measured brain activity during stimulation with speech stimuli in which prosody was manipulated in order to portray various emotional states. Studies using PET [25] or fMRI [26–28] generally emphasize the greater activation of the right temporal lobe and right inferior prefrontal cortex when attention is directed to emotional prosody, confirming earlier clinical work [29,30]. More recently, the neural bases of emotional perception in voice were studied outside the context of speech by using affective nonverbal vocalizations such as laughs, cries, groans and other more primitive vocal expressions of emotion. PET [31,32] and fMRI [33] studies have suggested the importance of structures such as the amygdala and anterior insula in processing vocal emotion.

Perception of identity information: speaker recognition

Relatively little is known about the neuronal bases of speaker perception and recognition. Several clinical studies have documented cases of brain-lesioned patients with a deficit in speaker discrimination or recognition [34–39]. These studies generally show that deficits in discriminating unfamiliar speakers or deficits in the recognition of familiar speakers ('phonagnosia') can be dissociated, but both seem to occur more often after lesions in the right hemisphere. Importantly, a double dissociation between speech perception and speaker recognition has been demonstrated by cases of preserved speech perception but impaired speaker recognition as well as cases of aphasia with normal voice recognition [35]. This supports a model of the organization of voice processing in which

Box 3. A model of voice perception

We propose to use Bruce and Young's model of face perception [86] as a framework for understanding the perceptual and cognitive processes involved in voice perception (see Figure I). After low-level analysis in subcortical nuclei and regions of primary auditory cortex (A1), vocal stimuli are further processed in a stage of 'structural encoding' – probably involving bilateral regions of the middle STS close to A1. Vocal information processing might then be dissociated in three functionally independent systems: (i) analysis of speech information, involving anterior and posterior STS as well as inferior prefrontal regions predominantly in the left hemisphere; (ii) analysis of vocal affective information, involving temporo-medial regions, anterior insula, and amygdala and inferior prefrontal regions predominantly in the right hemisphere; (iii) analysis of vocal identity, involving 'voice recognition units' – probably instantiated in regions of the right anterior STS – each activated by one of the voices known to the person, and a subsequent supra-modal stage of person recognition ('person identity nodes'). These three processing pathways are proposed to interact with homologous pathways in the face processing architecture, in a supra-modal stage of information processing.

This model predicts functional dissociations analogous to those observed or proposed for faces: patients should in principle be found with a dissociation between impaired processing of one type of vocal information and normal processing of the two other types of vocal information. Some of these dissociations have already been documented (e.g. double dissociation between receptive aphasia and phonagnosia) but others are still to be demonstrated. Standardized batteries of evaluation of voice perception abilities investigating speech as well as affect and identity perception would constitute a desirable tool.

Despite the proposed functional dissociation, the three pathways are clearly not wholly independent. During the normal processing of vocal information, cortical regions involved in processing the different types of vocal information are likely to interact to build increasingly abstract representations. It is only at the highest levels of the architecture that representations for one type of information would become independent of sources of variability related to other types of information. For example, the 'voice recognition units' are supposed to be activated by the voice of one individual regardless of the speech content or the emotional tone of the vocal input.

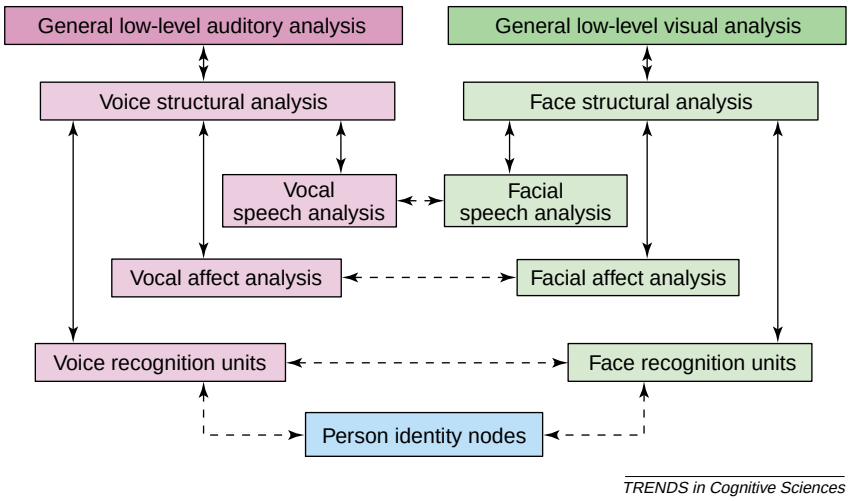

Figure I. A model of voice perception. The right-hand part of the figure is adapted from Bruce and Young's model of face perception [86]. The left-hand part proposes a similar functional organization for voice processing. Dashed arrows indicate multimodal interactions.

speech and identity information are processed in partially dissociated cortical regions (see Box 3).

Only few neuroimaging studies have investigated the perception of identity information. Imaimuzi and colleagues [40] were the first to use PET to examine patterns of cerebral activity induced by speaker identification. Subjects were scanned while performing a forced-choice identification of either the speaker or the emotion in non-emotional words pronounced by four actors with four different emotional tones. They found that in both hemispheres, the anterior temporal lobes were more active during speaker identification than during emotion identification [40]. In a subsequent study [41], the same group scanned normal volunteers with PET while they performed a familiar/unfamiliar decision task on voices from unknown persons and from their friends and relatives. A comparison task consisted of deciding whether the first phoneme of sentences pronounced by unfamiliar speakers was a vowel or a consonant. The results showed that several cortical regions, including the enthorinal cortex and the anterior part of the right temporal lobe, were more active during the voice familiarity task. Interestingly, the amount of activity in the right anterior temporal pole was found to correlate positively with the subjects' performance at a speaker identification task administered just after scanning.

Von Kriegstein and colleagues [42] used fMRI to measure brain activity during identification tasks directed either to the speaker's voice or to the verbal content of sentences in German. 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 (Figure 1), 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 or speech content was found to modulate activity in the STS regions. A convergent finding was obtained by Belin and Zatorre [43] in an fMRI study with an opposite design. There, two conditions shared a common passive listening task but blocks of vocal

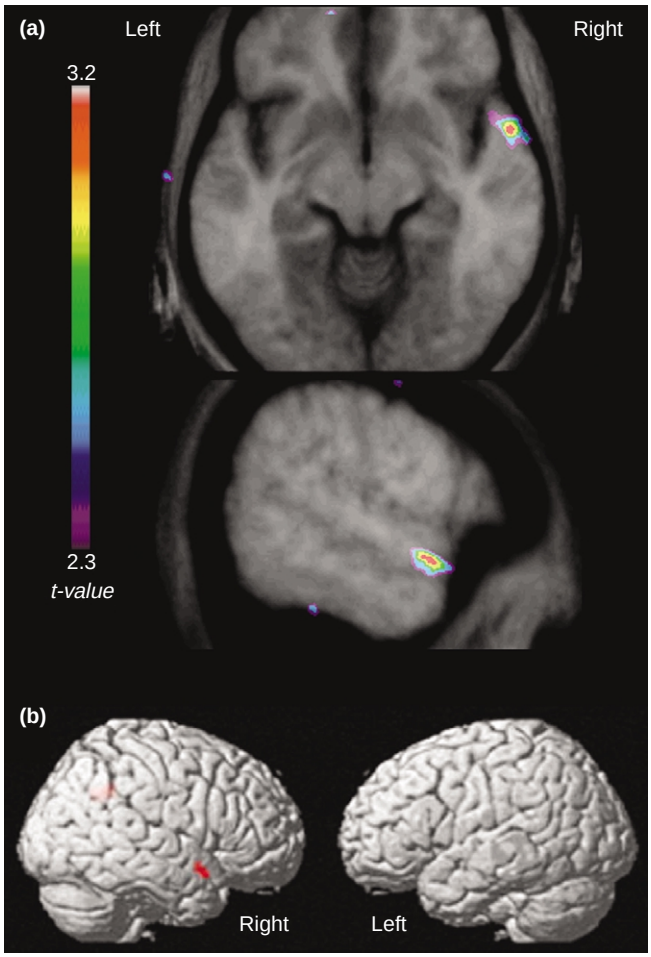

Figure 1. Cortical sensitivity to speaker's identity. **(a)** Cortical regions showing decrease in neuronal activity with repetition of the speaker's voice, shown in color-scale on axial (top) and sagittal (middle) slices through the subjects' mean anatomical image. (Reproduced with permission from [43].) **(b)** Rendering on a reconstructed cortex of regions showing greater activity (red) when attention is directed to speaker's voice compared with verbal content. (Reproduced with permission from [42]). Note similarity of outlined regions despite difference in paradigms.

stimulation were composed either of the same syllable spoken by 12 different speakers or of 12 syllables spoken by the same speaker – thus repeating either speaker or syllable. Only one region of the auditory cortex, in the right anterior STS, showed reduced activity when different syllables were pronounced by the same voice as compared with different voices saying the same syllable (see Figure 1). This reduced response to a same voice was interpreted as an adaptation response by neuronal populations sensitive to idiosyncratic acoustic features of a speaker's voice.

Thus, there is clear converging evidence for an important role of anterior temporal-lobe regions of the right hemisphere, particularly right anterior STS regions, in processing information related to speaker identity. This is consistent with recent models of the organization of the primate auditory cortex [44,45] in which a ventral 'What' pathway, homologous to the similar pathway in the visual system [46], would be specialized in recognition of auditory objects, and in particular, individual voices. Note, the STS is a long, heterogeneous structure: cyto-architectonic and connectivity studies in the rhesus monkey have demonstrated a division of the STS into several uni- or polymodal areas organized in a precise sequence of reciprocal connections with one another and with other regions of the cortex [47]. Thus, the various STS activations observed in neuroimaging studies probably correspond to several functionally distinct regions.

Face–voice integration

The integration of information from faces and voices is known to affect the processing of unimodal information. The McGurk effect [48], where incongruent facial and vocal phonetic information results in an intermediate percept, provides a dramatic illustration of this phenomenon. Neuroimaging studies have again mostly investigated the integration of phonetic information. The activation of a left posterior STS region during the perception of facial speech (lipreading) is generally observed, with increased activity for conditions of bimodal integration (e.g. [49,50]).

Less is known about the face–voice integration of identity information. Shah and colleagues [51] used fMRI to investigate the neural response to person familiarity through both visual and auditory modalities. Subjects were scanned while hearing voices or viewing faces belonging to persons either unknown or personally known. Processing of familiar faces and voices led to greater activation of the retrosplenial cortex, close to the precuneus activation found by von Kriegstein and colleagues for familiar voices [42].

Face and voice integration has also been studied in the context of affective information processing. Dolan and colleagues [52] investigated how the cortical response to facial expressions of fear was modulated by simultaneous presentation of voices in which prosody expressed happiness or fear. They found effects of integration in the left amygdala, in which activity was strongest when both faces and voices expressed fear, and smallest for incongruent happy–sad, facial–vocal pairs.

Together, these studies suggest that the integration of speech, identity and affective information from faces and voices involve different cortical regions; they again support the idea that the neuronal processing of these three types of information could be dissociated in similar ways for voices as for faces (see Box 3).

Are voices 'special'?

The debate on 'Is speech special ?', that is, on whether or not speech perception involves specialized, modular brain mechanisms has generated much literature, and still appears to be unresolved. Evidence for mechanisms uniquely involved in speech perception (e.g. [53,54]) seems balanced by findings that categorization of phonemes could be based on more general acoustic mechanisms also present in other animals [55,56]. We would like to suggest that this question can be extended to affective and identity processing in voice: does voice perception in general involve specialized mechanisms not used for other, non-vocal, sounds from the environment?

This question is close to another active debate in cognitive neuroscience: 'Are faces special?'. On the one hand, evidence from various sources suggest that face processing recruits mechanisms not normally involved in the processing of other objects. These results include: behavioral evidence for greater disruption of face than object processing by picture inversion (the 'face inversion effect' [57]); the observation of patients with impaired face recognition ('prosopagnosia') despite normal object recognition, or vice-versa [58]; electrophysiological recordings in the macaque STS showing cells with greater response to faces than other objects, although in small proportion [59]; evidence for an electrophysiological 'N170' negativity selective for faces over occipito-temporal cortex [60,61], and neuroimaging evidence for greater responses to faces than objects in discrete cortical regions [62–64]. On the other hand, some findings suggest that the above effects could be related to the greater difficulty of face discrimination, categorization and recognition tasks at a subordinate level, compared with other object categories where exemplars are much less similar to one another. For example, cortical regions considered as 'face-selective' also respond to a lesser degree to non-face objects, and they appear to be strongly activated by non-face stimuli such as birds or cars in subjects with expertise with those stimuli [65,66].

Voice-sensitivity in auditory cortex

Some of the arguments used in the face debate are now being transposed to the voice domain. In particular, recent results suggest the existence of cortical regions selective to sounds of voice. Belin and colleagues [67] used fMRI to measure brain activity during passive stimulation with a large variety of natural sounds grouped in blocks of either vocal or non-vocal sounds. Most parts of the auditory cortex responded similarly to vocal and to non-vocal sounds. However, in each subject discrete regions of auditory cortex were found that exhibited a greater response to the vocal sounds – no part of cortex showed the reverse pattern. These voice-sensitive cortical regions showed high selectivity, because their response to vocal sounds was significantly greater than to control sounds with similar low-level structure such as speech-envelope noise or scrambled voices. These regions were consistently located along the superior bank of the STS (see Figure 2). Interestingly, the group-average map of cortical activity suggested greater voice-selectivity in the right hemisphere, which could be related to the fact that half of the vocal stimuli consisted of non-speech vocalizations [67]. Only regions in the right anterior STS showed a greater response to non-speech vocal sounds than to their scrambled counterpart, providing additional evidence that these regions could be involved in processing paralinguistic information in voice [68].

Recent EEG and MEG studies also addressed the question of voice specificity. Levy and colleagues used ERPs to compare the response evoked by sung voices and tones played by different musical instruments. No difference between the voices and instruments was observed for the N1 component; however a 'voice-specific response' (VSR) could be observed, peaking at around 320 ms after stimulus onset and strongest on the right side. They suggested that this component, different from the 'novelty P300', might reflect allocation of attention related to the salience of voice stimuli [69,70]. Gunji and colleagues, in a MEG study using similar stimuli [71], also found no difference in the N1 evoked by voices and instruments respectively, but found a sustained field with greater source strength for the voice stimuli between 300 and 500 ms after onset. They did not observe, however, a magnetic counterpart of the VSR found with ERP, which they attributed to the radial orientation of the sources involved or to the movie viewing condition used in the MEG but not the ERP study. An open question, that could be addressed by future studies using high-density ERPs or combined fMRI and ERPs, is whether the VSR might originate in STS regions. It will also be interesting to test whether an earlier VSR, closer in time to the N170 observed for faces, will be observed with more varied sets of vocal and non-vocal stimuli such as those used in face perception research.

Thus, these studies provide evidence for cortical mechanisms activated by vocal stimuli more than by other non-vocal stimuli. But as with face-specific responses of visual cortex, these responses could be interpreted as a modular response to the category of vocal sounds, or as reflecting discriminations and categorization between highly similar

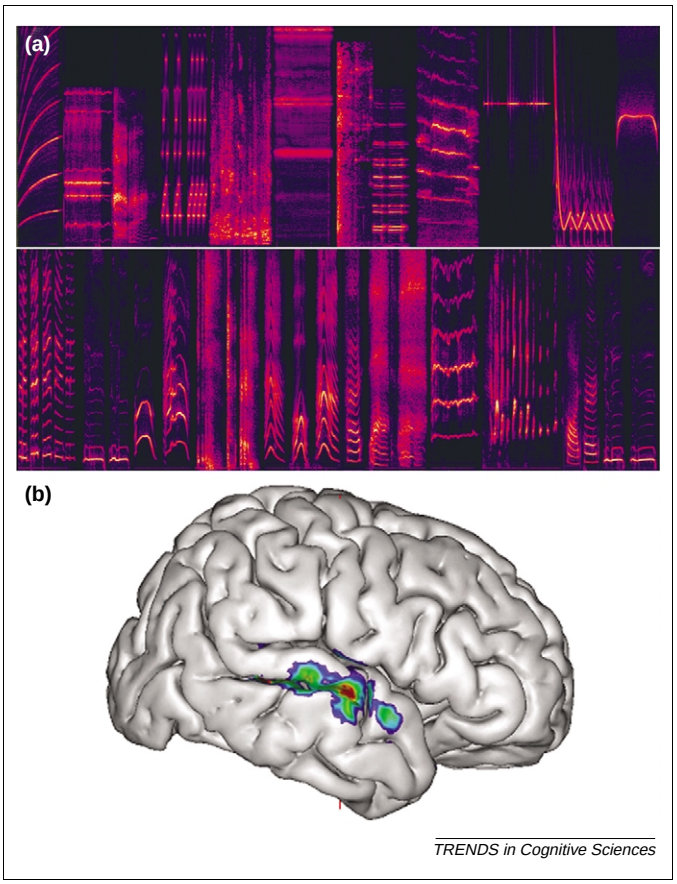

Figure 2. Voice-sensitive cortical activity. **(a)** Spectrograms (x: time; y: frequency; color indicates energy) of examples of non-vocal (top) and vocal (bottom) sounds used in [67]. Note their similar apparent complexity. **(b)** Rendering on a reconstructed cortex of cortical regions showing greater response to vocal compared with non-vocal sounds in 8 subjects, located in the anterior part of the STS.

Box 4. Questions for future research

- Would STS activations also be observed for expert categorization at a subordinate level of other sound categories ?
- Are voices more 'attention-grabbing' than other sounds, whether they contain speech or not?
- To what extent does our voice perception system allow us to extract identity and affective information from vocalizations from other species, such as cats and dogs?
- What are the perceptual primitives of voices? Can voices be represented with a small number of independent dimensions?
- Are there people with 'congenital phonagnosia', that is, a developmental inability to recognize voices? (Landis, pers. commun.).

exemplars of a sound category, and might therefore also be present for expert categorization of other sounds (see Box 4).

Conclusion

Recent neuroimaging studies suggest that different cortical regions are involved in processing different types of vocal information, with an important role of regions along the STS. They suggest a model of functional organization similar to those proposed for face perception, where linguistic, affective, and identity information are processed in partially segregated cortical pathways. Further work – possibly using paradigms and tools inspired from the face processing literature – will be necessary to test the predictions of this model, and confirm or refute the analogy to face perception.

Acknowledgements

We thank Robert Zatorre, Isabelle Peretz, Frédéric Gosselin and Marie-Hélène Grosbras for useful comments on the manuscript. This work is supported by grants from National Sciences and Engineering Research Council of Canada and Fonds Québecois de Recherche sur la Nature et les Technologies.

References

1 Hauser, M. (1996) *The Evolution of Communication*, MIT Press
2 Fitch, W.T. (2000) The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267
3 Ellis, A.W. (1989) Neuro-cognitive processing of faces and voices. In *Handbook of Research on Face Processing* (Young, A.W. and Ellis, H.D., eds), pp. 207–215, Elsevier
4 Lass, N.J. *et al.* (1976) Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* 59, 675–678
5 Mullennix, J.W. *et al.* (1995) The perceptual representation of voice gender. *J. Acoust. Soc. Am.* 98, 3080–3095
6 Hartman, D.E. and Danahuer, J.L. (1976) Perceptual features of speech for males in four perceived age decades. *J. Acoust. Soc. Am.* 59, 713–715
7 Linville, S.E. (1996) The sound of senescence. *J. Voice* 10, 190–200
8 Kreiman, J. (1997) Listening to voices: theory and practice in voice perception research. In *Talker Variability in Speech Research* (Johnson, K. and Mullenix, J., eds), pp. 85–108, Academic Press
9 Papcun, G. *et al.* (1989) Long-term memory for unfamiliar voices. *J. Acoust. Soc. Am.* 85, 913–925
10 Schweinberger, S.R. *et al.* (1997) Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *J. Speech Lang. Hear. Res.* 40, 453–463
11 Monrad-Krohn, G.H. (1963) The third element of speech: prosody and its disorders. In *Problems of Dynamic Neurology* (Halpern, L., ed.), pp. 101–117, Hebrew University Press
12 Scherer, K.R. (1995) Expression of emotion in voice and music. *J. Voice* 9, 235–248
13 Klatt, D.H. (1980) Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971–995
14 Remez, R.E. *et al.* (1981) Speech perception without traditional speech cues. *Science* 212, 947–950
15 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
16 Remez, R.E. *et al.* (1997) Talker identification based on phonetic information. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 651–666
17 Hickok, G. and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138
18 Zatorre, R.J. and Binder, J.R. (2000) Functional and structural imaging of the human auditory system. In *Brain Mapping: The Systems*, pp. 365–402, Academic Press
19 Samson, Y. *et al.* (2001) Auditory perception and language: functional imaging of speech sensitive auditory cortex. *Rev. Neurol. (Paris)* 157, 837–846
20 Scott, S.K. and Johnsrude, I.S. (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107
21 Binder, J.R. *et al.* (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528
22 Scott, S.K. *et al.* (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406
23 Davis, M.H. and Johnsrude, I.S. (2003) Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431
24 Giraud, A.L. *et al.* Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb. Cortex* (in press)
25 George, M.S. *et al.* (1996) Understanding emotional prosody activates right hemisphere region. *Arch. Neurol.* 53, 665–670
26 Buchanan, T.W. *et al.* (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res. Cogn. Brain Res.* 9, 227–238
27 Wildgruber, D. *et al.* (2002) Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *NeuroImage* 15, 856–869
28 Mitchell, R.L. *et al.* (2003) The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41, 1410–1421
29 Ross, E.D. (1981) The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Arch. Neurol.* 38, 561–569
30 Heilman, K.M. *et al.* (1984) Comprehension of affective and non-affective prosody. *Neurology* 34, 917–921
31 Phillips, M.L. *et al.* (1998) Neural responses to facial and vocal expressions of fear and disgust. *Proc. R. Soc. Lond. B. Biol. Sci.* 265, 1809–1817
32 Morris, J.S. *et al.* (1999) Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia* 37, 1155–1163
33 Sander, K. and Scheich, H. (2001) Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res. Cogn. Brain Res.* 12, 181–198
34 Assal, G. *et al.* (1976) Discrimination des voix lors des lesions du cortex cerebral. *Schweiz. Arch. Neurol. Neurochir. Psychiatr.* 119, 307–315
35 Assal, G. *et al.* (1981) Asymétrie cérébrale et reconnaissance de la voix. *Rev. Neurol. (Paris)* 137, 255–268
36 Van Lancker, D.R. and Canter, G.J. (1982) Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn.* 1, 185–195
37 Van Lancker, D. and Kreiman, J. (1987) Voice discrimination and recognition are separate abilities. *Neuropsychologia* 25, 829–834
38 Peretz, I. *et al.* (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117, 1283–1301
39 Neuner, F. and Schweinberger, S.R. (2000) Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn.* 44, 342–366
40 Imaizumi, S. *et al.* (1997) Vocal identification of speaker and emotion activates different brain regions. *NeuroReport* 8, 2809–2812
41 Nakamura, K. *et al.* (2001) Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* 39, 1047–1054

42 von Kriegstein, K. *et al.* (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* 17, 48–55
43 Belin, P. and Zatorre, R.J. (2003) Adaptation to speaker's voice in right anterior temporal-lobe. *NeuroReport* 14, 2105–2109
44 Kaas, J.H. and Hackett, T.A. (1999) 'What' and 'where' processing in auditory cortex. *Nat. Neurosci.* 2, 1045–1047
45 Rauschecker, J.P. and Tian, B. (2000) Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11800–11806
46 Ungerleider, L.G. and Haxby, J.V. (1994) 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165
47 Seltzer, B. and Pandya, D.N. (1989) Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* 290, 451–471
48 McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 746–748
49 Wright, T.M. *et al.* (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043
50 Sekiyama, K. *et al.* (2003) Auditory-visual speech perception examined by fMRI and PET. *Neurosci. Res.* 47, 277–287
51 Shah, N.J. *et al.* (2001) The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain* 124, 804–815
52 Dolan, R.J. *et al.* (2001) Crossmodal binding of fear in voice and face. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10006–10010
53 Liberman, A.M. and Mattingly, I.G. (1989) A specialization for speech perception. *Science* 243, 489–494
54 Liberman, A.M. and Whalen, D.H. (2000) On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196
55 Ramus, F. *et al.* (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351
56 Kluender, K.R. *et al.* (1987) Japanese quail can learn phonetic categories. *Science* 237, 1195–1197
57 Yin, R.K. (1969) Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145
58 Farah, M.J. (1996) Is face recognition 'special'? Evidence from neuropsychology. *Behav. Brain Res.* 76, 181–189
59 Perrett, D.I. *et al.* (1992) Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 23–30
60 George, N. *et al.* (1996) Brain events related to normal and moderately scrambled faces. *Brain Res. Cogn. Brain Res.* 4, 65–76
61 McCarthy, G. *et al.* (1999) Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb. Cortex* 9, 431–444
62 Puce, A. *et al.* (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74, 1192–1199
63 Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311
64 Haxby, J.V. *et al.* (2000) The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233
65 Gauthier, I. *et al.* (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197
66 Tarr, M.J. and Cheng, Y.D. (2003) Learning to see faces and objects. *Trends Cogn. Sci.* 7, 23–30
67 Belin, P. *et al.* (2000) Voice-selective areas in human auditory cortex. *Nature* 403, 309–312
68 Belin, P. *et al.* (2002) Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* 13, 17–26
69 Levy, D.A. *et al.* (2001) Processing specificity for human voice stimuli: electrophysiological evidence. *NeuroReport* 12, 2653–2657
70 Levy, D.A. *et al.* (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology* 40, 291–305
71 Gunji, A. *et al.* (2003) Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci. Lett.* 348, 13–16
72 DeCasper, A.J. and Fifer, W.P. (1980) Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176
73 Ockleford, E.M. *et al.* (1988) Responses of neonates to parents' and others' voices. *Early Hum. Dev.* 18, 27–36
74 Fifer, W.P. and Moon, C.M. (1994) The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatr. Suppl.* 397, 86–93
75 Kisilevsky, B.S. *et al.* (2003) Effects of experience on fetal voice recognition. *Psychol. Sci.* 14, 220–224
76 Rendall, D. *et al.* (1998) The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* 103, 602–614
77 Rendall, D. *et al.* (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 51, 1007–1015
78 Insley, S.J. (2000) Long-term vocal recognition in the northern fur seal. *Nature* 406, 404–405
79 Charrier, I. *et al.* (2001) Mother's voice recognition by seal pups. *Nature* 412, 873
80 Fant, G. (1960) *Acoustic Theory of Speech Production*, Mouton
81 Titze, I.R. (1989) Physiologic and acoustic differences between malen and female voices. *J. Acoust. Soc. Am.* 85, 1699–1707
82 Klatt, D.H. and Klatt, L.C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820–857
83 Hanson, H.M. and Chuang, E.S. (1999) Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *J. Acoust. Soc. Am.* 106, 1064–1077
84 Sachs, J. *et al.* (1973) Anatomical and cultural determinants of male and female speech. In *Language Attitudes: Current Trends and Prospects* (Shuy, R.W. and Fasold, R.W., eds), pp. 74–84, Gerorgetown University Press
85 Fitch, W.T. and Giedd, J. (1999) Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522
86 Bruce, V. and Young, A. (1986) Understanding face recognition. *Br. J. Psychol.* 77, 305–327

TICS Book Reviews

We aim to review the best of the new books in Cognitive Science in *TICS.*

Our book reviews are styled like short opinion pieces or mini-reviews of a subject area, with the aim of providing readers with more in-depth analysis. As well as critically appraising important new texts in the field, the authors of reviews will be encouraged to use the book as a framework for wider discussion.

Publishers: – monographs and edited volumes (but not undergraduate textbooks) will be considered for review in *Trends in Cognitive Sciences*. Please send us advance e-mails of forthcoming titles and copies of books you would like to be reviewed to the Book Review Editor at the address below.

Trends in Cognitive Sciences
Elsevier London, 84 Theobald's Road, London WC1X 8RR, UK.
e-mail: tics@current-trends.com