



HAL
open science

Hemispheric association and dissociation of voice and speech information processing in stroke

Anna B Jones, Andrew J Farrall, Pascal Belin, Cyril Pernet

► **To cite this version:**

Anna B Jones, Andrew J Farrall, Pascal Belin, Cyril Pernet. Hemispheric association and dissociation of voice and speech information processing in stroke. *Cortex*, 2015. hal-01997402

HAL Id: hal-01997402

<https://amu.hal.science/hal-01997402>

Submitted on 29 Jan 2019

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

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



Distributed under a Creative Commons Attribution 4.0 International License



Research report

Hemispheric association and dissociation of voice and speech information processing in stroke



Anna B. Jones^{a,b}, Andrew J. Farrall^{a,b}, Pascal Belin^{c,d} and Cyril R. Pernet^{a,b,*}

^a Brain Research Imaging Centre, The University of Edinburgh, UK

^b Centre for Clinical Brain Sciences, The University of Edinburgh, UK

^c Institute of Neuroscience and Psychology, University of Glasgow, UK

^d Institut des Neurosciences de La Timone, UMR 7289, CNRS & Université Aix-Marseille, France

ARTICLE INFO

Article history:

Received 9 January 2015

Reviewed 21 April 2015

Revised 22 May 2015

Accepted 6 July 2015

Action editor Peter Garrard

Published online 16 July 2015

Keywords:

Voice gender perception

Phoneme perception

Hemispheric dissociation

Stroke

Aphasia

ABSTRACT

As we listen to someone speaking, we extract both linguistic and non-linguistic information. Knowing how these two sets of information are processed in the brain is fundamental for the general understanding of social communication, speech recognition and therapy of language impairments. We investigated the pattern of performances in phoneme versus gender categorization in left and right hemisphere stroke patients, and found an anatomic-functional dissociation in the right frontal cortex, establishing a new syndrome in voice discrimination abilities. In addition, phoneme and gender performances were most often associated than dissociated in the left hemisphere patients, suggesting a common neural underpinnings.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Speech perception is often seen as special (Liberman & Mattingly, 1989) because localized brain injury can elicit specific language impairments such as aphasia, and because healthy individuals are extremely efficient at categorizing phonemes and syllables despite large variations in the stimulus spectral patterns (Liberman, Delattre, & Cooper, 1952). To

achieve high performance levels, it has been hypothesized that voice information (talker specific information) is extracted along with the speech signal, and then stripped away to access (invariant) phonemic content: a process known as ‘speaker normalization’. This hypothesis is however challenged because general auditory learning mechanisms are capable of explaining category formation in the absence of invariant acoustic information. Birds can learn speech consonant categories with no obvious acoustic invariant cue (Kluender, Diehl,

* Corresponding author. Centre for Clinical Brain Sciences (CCBS), The University of Edinburgh, Chancellor’s Building, Room GU426D, 49 Little France Crescent, Edinburgh EH16 4SB, UK.

E-mail address: cyril.pernet@ed.ac.uk (C.R. Pernet).

<http://dx.doi.org/10.1016/j.cortex.2015.07.004>

0010-9452/© 2015 Elsevier Ltd. All rights reserved.

& Killeen, 1987) and human listeners can readily learn non-speech categories that are similarly structured (Wade & Holt, 2005). In addition, several studies showed that talker variability influences speech perception. For instance the literature describes increased memory for words spoken by familiar voices, compared to non-familiar voices (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993), and similarly enhanced discrimination of, and memory for, (non-familiar) speakers of our own language compared to speakers of another language (Language Familiarity Effect – Perrachione & Wong, 2007) even in the absence of intelligibility (Fleming, Giordano, Caldara, & Belin, 2014). Most of these studies do not, however, specifically address the issue of phoneme perception, and thus acoustical regularities coming from multiple levels are at play.

Like speech perception, voice perception is often considered special (Belin, 2006; Belin, Fecteau, & Bedard, 2004; Scott, 2008). Humans easily recognize different voices, and this ability is of considerable social importance. Voice-selective areas have been demonstrated in the human brain (Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Pernet et al., 2015), localized bilaterally along the upper bank (middle and anterior) of the Superior Temporal Sulcus (STS) (Alho et al., 2006; Belin, et al., 2002), and also in the inferior and orbitofrontal cortex (Charest, Pernet, Latinus, Crabbe, & Belin, 2012; Fecteau, Armony, Joanete, & Belin, 2005) as well as surrounding insular cortex (Johnstone, van Reekum, Oakes, & Davidson, 2006; Rama et al., 2004). This neural selectivity for voice has also been established in other mammals, in particular primates (Johnstone et al., 2006; Rama et al., 2004) and more recently dogs (Andics, Gacsi, Farago, Kis, & Miklosi, 2014). Given the presence of con-specific voice neural selectivity in these animals, we can establish that neural voice selectivity is an old evolutionary feature (~100 million years for a common ancestor between humans and dogs, ~25 million years for a common ancestor between macaques and human), preceding the appearance of speech (~5 to 2 million years ago for proto-language and ~150,000 to 300,000 years ago for speech – Perreault & Mathew, 2012). Following psycholinguistic studies suggesting that phonetic attributes are an intrinsic component of the perception and identification of individuals, i.e., the recognition of voices (Remez, Fellowes, & Rubin, 1997), it is possible that some brain regions dedicated to speech co-opted neurons already involved in con-specific voice processing.

In this study, we investigated how phoneme and talker information processing relate to each other, by comparing performances of right fronto-temporal (non-aphasic), left fronto-temporal aphasic and left fronto-temporal non-aphasic stroke patients. Each participant categorized sounds from pitch equalized morphed continua as being male–female or/pa/–/ta/(Pernet, Belin, & Jones, 2014). Stimuli were the same in both tasks, and participants had therefore to discard talker specific or phoneme specific information depending on the task at hand. Given the importance of the right STS (Besteplemeyer, P. Belin, & Grosbras, 2011) and right Middle and Inferior Frontal Gyrus (MFG-IFG) (Charest et al., 2012) in talker information processing, we hypothesized that right hemisphere patients will show a dissociation between the two tasks. In contrast, following our hypothesis of co-optation of voice selective neurons in phoneme processing,

we hypothesized that left hemispheric aphasic patients will not show such dissociation, while non-aphasic patient could be impaired for voice but not phoneme.

2. Materials and methods

The experiment used (program and stimuli) is freely available from <http://dx.doi.org/10.6084/m9.figshare.1287284>. It runs under Matlab with the psychophysical toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). The behavioral data and scripts used to analyze the data are available from <http://dx.doi.org/10.6084/m9.figshare.1287262>. The imaging radiological analysis is also available with the behavioral data. CT or MRI scans could not be shared as they belong to the UK National Health Service (NHS) and not to the research team. The study was approved by the NHS Lothian South East Scotland Research Ethics Committee 01 (REC reference number: 11/SS/0055) and NHS Lothian Research and Development (R&D project number: 2011/W/NEU/09).

2.1. Participants

Twenty-five stroke patients (14 males, 11 females) with a median age of 69 years (min 39, max 85) were recruited into this study. At the time of testing, all patients were at the chronic stage (median time between stroke and testing 90 ± 17 days). Participants were recruited as inpatients and outpatients from Lothian NHS hospitals via stroke physicians and Speech & Language Therapists between 10 and 60 weeks post-stroke with the sole inclusion criterion of a stroke affecting perisylvian tissues (Supplementary Table 1). Exclusion criteria were the presence of a previous stroke and/or English not being the participant's first language.

All patients were tested for their mood (Visual Analogue Self Esteem Scale – VASES (Brumfitt & Sheeran, 1999) and The Depression Intensity Scale Circles – DISCS (Turner-Stokes, Kalmus, Hirani, & Clegg, 2005) and language abilities (Western Aphasia Battery – WAB, Shewan & Kertesz, 1980). No patient had language deficits in the group with right hemisphere lesions ($N = 9$, 5 males and 4 females, WAB median score 98.8), and 10 out of 16 patients showed signs of aphasia in the left hemisphere group ($N = 10$, 5 males and 5 females, WAB median score 49.5 for aphasics vs $N = 6$, 4 males and 2 females, WAB median score 99.4 for non-aphasics – percentile bootstrap difference 49.8 [15 80] $p = 0$). Kruskal-Wallis ANOVA showed that groups did not differ in terms of median age [$\chi^2(2,22) = 4.58$ $p = .1$], in median time delay between stroke and testing [$\chi^2(2,22) = 1.68$ $p = .43$] or depression scores [$\chi^2(2,22) = 4$ $p = .13$ for VASES and $\chi^2(2,22) = 2.19$ $p = .33$ for DISCS]. All demographic and test results are summarized in Supplementary Table 2.

2.2. Paradigm

The experiment was identical to Pernet et al. (2014), except that only pitch equalized stimuli were used. Participants carried out two 2 alternative forced choice identification tasks: voice gender (male vs female) and phoneme (/pa/vs/ta/), and responded by button press on a keyboard. For each task, the

same two continua of morphed sounds were used: the 1st continuum going from a Male-/pa/to a Female-/ta/and the 2nd continuum with the same speakers going from a Male-/ta/to Female-/pa/. Morphs were generated by steps of 10% giving for the 1st continuum, 100% Male-/pa/, 90% Male-/pa/with 10% Female-/ta/, 80% Male-/pa/with 20% Female-/ta/, etc. until 100% Female-/ta/and for the 2nd continuum, 100% Male-/ta/, 90% Male-/ta/with 10% Female-/pa/, 80% Male-/ta/with 20% Female-/pa/, etc. until 100% Female-/pa/. This design allowed investigation of the effect of the task while controlling for the general acoustic characteristics of the stimuli, since the same stimuli were used in both tasks. Participants heard each stimulus in pseudo-random order six times each, for a total of 132 stimuli (2 continua * 11 steps * 6 trials) per task. Eighteen different continua of stimuli were generated from 6 different speakers (3 males and 3 females pronouncing/pa/and/ta/) and randomly assigned to participants. Task order and key orientation were counterbalanced between participants. Between each task, an interfering tone discrimination task was also performed. Participants heard pure tones of various frequencies corresponding to the male and the female ranges and had to tell if 2 consecutive sounds were the same or different. The task followed a 2 down, 1 up step-wise procedure (H. Levitt, 1971) equating participants performances when the staircase has ended (70.71% percent correct). This task was primarily designed to minimise the influence of one categorization task on the other, but also allowed control for basic auditory impairments. No significant differences were observed between groups on this task [Kruskal-Wallis ANOVA, $\chi^2(2,22) = 1.93$ $p = .38$, [Supplementary Table 2](#)].

2.3. Behavioral classification

To assess the independence of phoneme and gender categorization tasks, behavioral performances were binarized as impaired versus unimpaired. For each subject, response proportion curves (percentage of female response or percentage of /ta/responses) were obtained by averaging repeated trials from the different continua (Fig. 1). Each participant was then classified based on his or her ability to perform outside chance level, at least one time for the first 3 stimuli and at least one time for the last 3 stimuli along the sound continua. This implies that if a participant answered correctly for a least one of the initial stimuli (100% or 90% or 80% male or/pa/) and one of the final stimuli (100% or 90% or 80% female or/ta/), he or she was considered unimpaired (see [Supplementary material 1](#) for repeated analyses using an incremental classification criterion). In normal healthy participants, this is achieved very easily. Taking the data from [Pernet et al. \(2014\)](#), 100% of controls ($N = 18$, 9 males, 9 females) were unimpaired (Fig. 2). From the resulting classification in patients, the independence between the phoneme and gender categorization performance was tested for each group using a McNemar test with exact central probability (Fay, 2010).

2.4. Lesion to symptom mapping

Each participant received one or more CT or MRI scan from the National Health Service, and the scan the closest in time to the subjects' participation in this study was considered. One-to-

one mapping was computed using a McNemar test with exact central probability (Fay, 2010) between behavioral deficits (impaired/non-impaired) and 6 Regions Of Interest (ROI – lesioned/non-lesioned). Significance was considered at $\alpha = .0083$ i.e., Bonferroni corrected for the 6 ROI. The ROI classification was performed by an expert neuro-radiologist (AF) with 12 years of experience, and following a detailed protocol¹ adapted from the International Stroke Trial III (Whiteley, Lindsey, Wardlaw, & Sandercock, 2006). Regions of Interest considered were the middle/inferior frontal gyrus (involved in gender categorization, Fecteau, et al., 2005; Charest et al., 2012), Heschl's gyrus (also called Transverse Temporal gyrus – involved in language learning and spectral/pitch information processing, Warrier et al., 2009), the superior temporal gyrus (anterior/posterior, involved in general auditory processes but also voice perception, Belin, et al., 2004), the insula (involved in central auditory functions, in particular temporal resolution and sequencing, Bamiau et al., 2006) and the amygdala–hippocampal complex (involved in memory and emotional voice perception, Johnstone et al., 2006; Rama et al., 2004).

2.5. Behavioral quantitative analyses

Each participant ability to distinguish between male–female and/pa/–/ta/stimuli was investigated using signal detection theory (Macmillan & Creelman, 2005), computing the perceptual distance d' (i) between successive pairs of stimuli along each continuum, and (ii) between the extreme stimuli of each continuum (global d' – the distance between 100% male or/pa/ and 100% female or/ta/). We speak of categorical perception when a continuum of stimuli is divided by a boundary, such that stimuli on one side of the boundary are perceived as belonging to one category, whereas stimuli on the other side are perceived as belonging to another category, and perceptual distances within a category are low. Here we investigated categorical perception in each group and task by testing if d' values differed from 0 (bootstrap-t test with Bonferroni correction – Wilcox, 2012) for each pair of stimuli along the continuum, thus delineating a perceptual boundary. Finally, differences between gender and phoneme categorization performances were tested comparing global d' values within groups (percentile bootstrap of the difference with adjustment for multiple testing) and relative to the healthy participants from [Pernet et al. \(2014\)](#) (mean difference between groups, with adjustment for multiple comparisons based on the maximum statistics – Wilcox, 2012).

3. Results

3.1. Behavioral classification

Of the 25 patients recruited, 19 showed at least one categorization deficit (Table 1), which we defined as the inability to perform above chance for at least one of three extreme stimuli (100%, 90% or 80% male/pa, female/pa, male/ta or female/ta).

¹ http://www.bric.ed.ac.uk/documents/CT_MR_for_voice_perception.pdf.

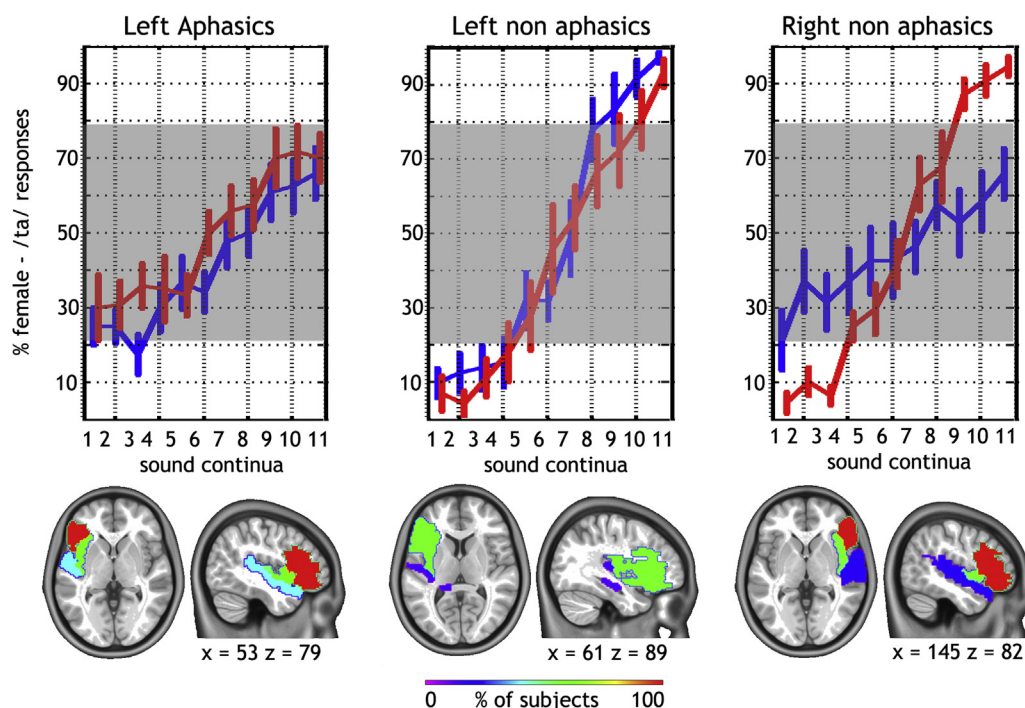


Fig. 1 – Average response proportion curves for each group of subjects (red lines represent phoneme task; blue lines represent gender task, bars represent bootstrap standard errors, shaded areas indicate chance level from 21.09% to 78.91%). At the bottom is shown the distribution of lesions per ROI for each group superimposed onto the MNI 152 template.

As hypothesized, we observed a significant dissociation in right fronto-temporal patients with impaired voice gender categorization versus intact phonological categorization (8 out of 9 patients, $\chi^2 = 7$, $p = 0$, $\Phi = \text{inf}$). In left fronto-temporal patients, we found no dissociation, for both the aphasic and the non-aphasic groups. Aphasic patients tended to show both phonological and voice gender categorization deficits ($\chi^2 = .2$, $p = 1$, $\Phi = .6$) and non-aphasic patients tended to performed normally in both tasks ($\chi^2 = 0$, $p = .5$, $\Phi = 1$ – Fig. 1). The same association/dissociation patterns were observed when varying the categorization deficit criteria (Supplementary material 1).

3.2. Lesion symptom mapping

One-to-one mapping between the behavioral classifications (impaired/unimpaired) and region-of-interests (lesioned/not-lesioned) showed that gender categorization impairments are associated with right frontal lesions ($\chi^2 = 8$, $p = .0078$ – see Fig. 2 for details). No other ROI shows significant results.

3.3. Behavioral quantitative analyses

Analyses of perceptual distances (d') between successive pairs of items along continua revealed that none of the patient groups had increased perceptual distances for ambiguous items (Fig. 3), contrary to healthy subjects as shown in Pernet et al. (2014). This result indicates a generalized reduction in categorical boundaries following stroke. Further analyses on global perceived distances (i.e., the d' computed between 100%

male and 100% female or 100%/pa/and 100%/ta/stimuli) showed that, compared to controls, right fronto-temporal patients had a lower global d' in the gender categorization task only (difference controls vs patients = 1.3, adjusted 95% CI [.8 1.8] for Male/Female, difference controls vs patients = .06, adjusted 95% CI [–.2 .3] for Pa/Ta). For left fronto-temporal aphasic patients, global d' was lower than in controls for both Male/Female stimuli (difference controls vs patients = 1.4 adjusted 95% CI [.5 2.2]) and pa/ta stimuli (difference controls vs patients = 1.4, adjusted 95% CI [.4 2.4]). For left fronto-temporal non-aphasic patients, global d' did not differ from controls (difference controls vs patients = .11 adjusted 95% [–.27 .49] for Male/Female, difference controls vs patients = .17, adjusted 95% CI [–.49 .83] for Pa/Ta). Finally, looking at within group differences, whilst right fronto-temporal patients show a significant difference between tasks, no significant differences were observed for both left fronto-temporal groups (Fig. 1, Table 2), which concurs with classification results.

4. Discussion

Based on functional MRI (Belin, et al., 2002, 2000; Charest et al., 2012) and Transcranial Magnetic Stimulation (Besteleyen, et al., 2011) results observed in healthy volunteers, we hypothesized that patients with right fronto-temporal stroke will show a deficit in gender categorization but intact phonological performances. Our results using both a qualitative (classification of percentage of responses) and a

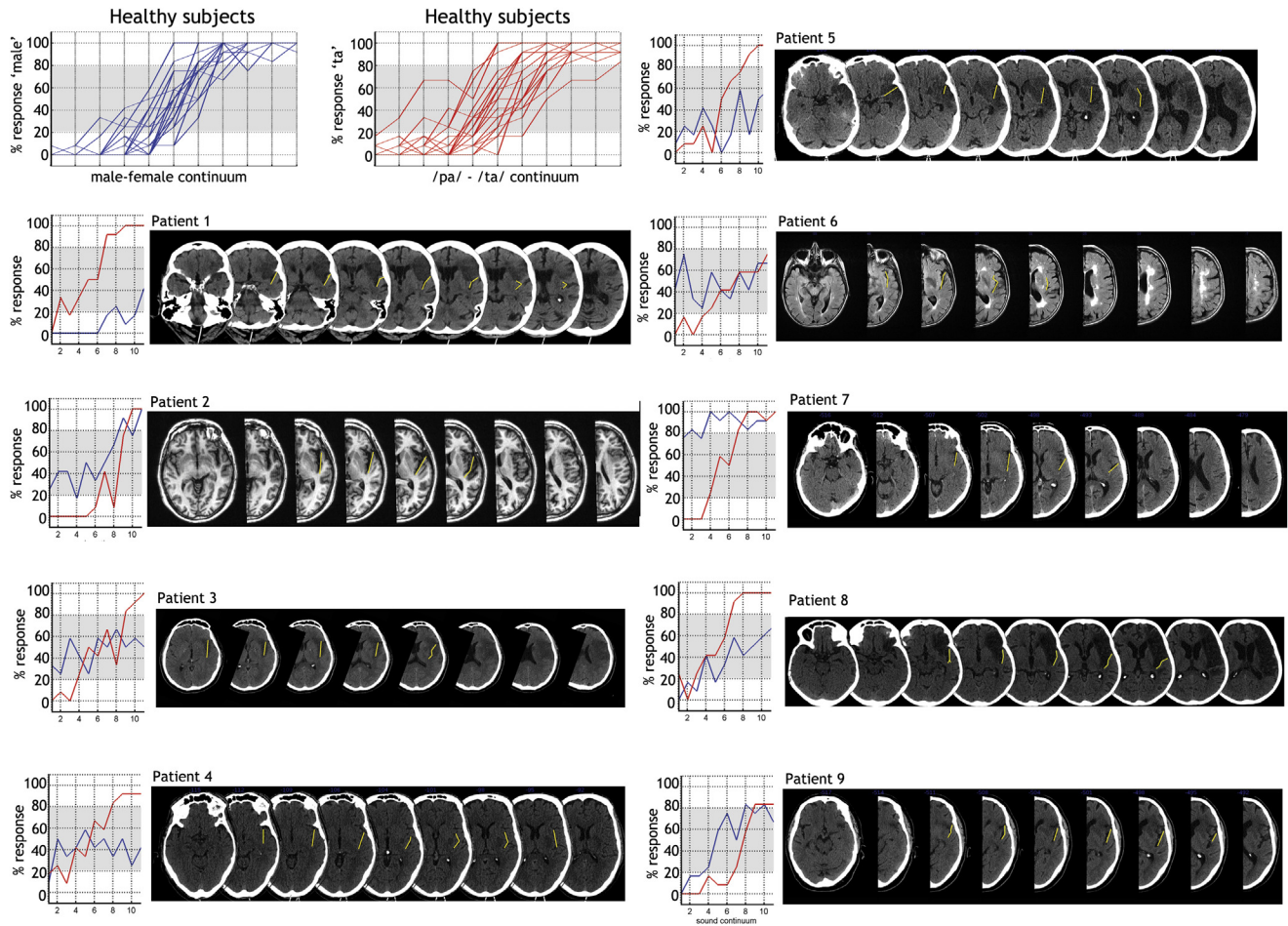


Fig. 2 – At the top left is shown the response proportion curves in normal healthy subjects (Pernet et al., 2014), illustrating that the classification works for well with controls scoring at 100% correct. Below are the same response proportion curves but for each right hemisphere patient, associated with axial slices showing the lesion in relation to the Sylvian fissure (highlighted in yellow). Importantly, the STG is intact in 7 out of 9 cases (patients 3 and 5 having both frontal and temporal lesions), suggesting that this region is not critical for gender/voice categorization.

quantitative (signal detection theory) approach confirmed this hypothesis. To our knowledge this is the first time that such deficit has been described in the literature. All of the patients presenting right frontal lesions showed a deficit in voice categorization, thus demonstrating a significant brain/behavior association. Previous studies have reported cases of phonagnosia (Van Lancker & Canter, 1982; Van Lancker, Cummings, Kreiman, & Dobkin, 1988; Van Lancker, Kreiman, & Cummings, 1989), in which patients could not recognize

familiar voices; but this deficit was associated with right parietal lesions. When discrimination of unfamiliar voices was tested, deficits were associated with temporal (left or right) lesions although some evidence also exists for voice deficit during fronto-temporal degeneration (Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010; Hailstone, Crutch, & Warren, 2010). What remains unclear is (1) if the deficit is specific to gender categorization or if it also relates to identity and (2) what is the role of the right IFG. On one hand, studies

Table 1 – Contingency table showing the classification of patients (per group) for the gender and the phoneme tasks.

		Gender task		
		Unimpaired	Impaired	
Phoneme task	Left fronto-temporal aphasics	Unimpaired	1	2
		Impaired	3	4
	Left fronto-temporal Non-aphasics	Unimpaired	4	1
		Impaired	1	0
Right fronto-temporal		Unimpaired	1	7
		Impaired	0	1

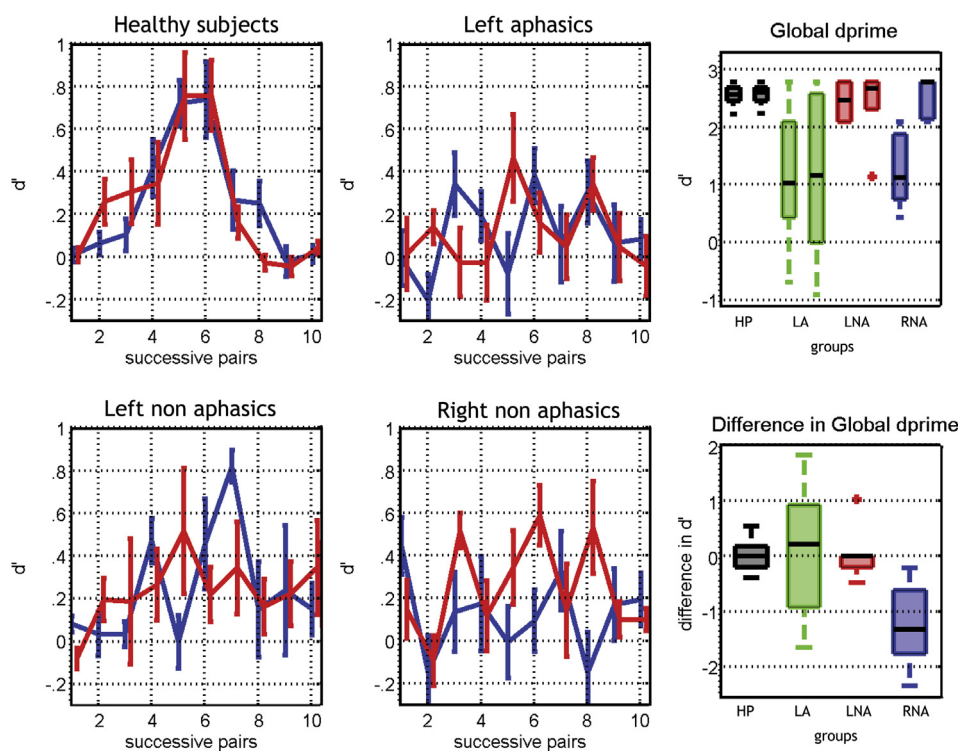


Fig. 3 – Perceptual and global perceptual distances (d'). Curves show d' values in the gender categorization task (blue) and the phoneme categorization task (red). Bars represent bootstrap standard errors. Box plots show the global d' and the difference in global d' between the two tasks for control healthy participants (HP – black), left aphasic (LA – green), left non-aphasic (LNA – red) and right non-aphasic (RNA – blue) patients. Boxes represent the interquartile range with the median, and whiskers show the 1.5*interquartile distance with outliers marked beyond that points.

Table 2 – Global perceived distance (d') computed for each group with within group differences, adjusted 95% confidence intervals and p values.

	Controls	Left fronto-temporal aphasics	Left fronto-temporal non aphasics	Right fronto-temporal
Distance Male/Female	2.54	1.12	2.43	1.23
Distance Pa/Ta	2.55	1.1	2.17	2.49
Difference and adjusted 95% CI	.01 [-.1 .1] $p = .872$.019 [.9 .8] $p = .958$.054 [-.4 .6] $p = .86$	1.25 [1.8 .7] $p = 0$

comparing attention to voice versus speech found voice specific effects over the right STS (Von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003), and this effect has been related to speaker identity (Schall, Kiebel, Maess, & von Kriegstein, 2014). On the other hand, using a continuous carry over-over design allowing acoustic from perceived distances to be distinguished during gender categorization, Charest et al., (2012) showed that the STS processes gender (and thus identity as well) related acoustic information whilst the right IFG is involved in perceived gender related distances. In most studies, stimuli are pitch equalized. We previously showed that equalizing pitch does not influence performance or RT (Pernet & Belin, 2012; Pernet et al., 2014); thus making timbre and consequently spectro-temporal analysis a key element in gender categorization. There is no doubt that the patients in our study found the gender categorization task difficult, even when the STS was intact (Fig. 2 – patients 1, 2, 4,

6, 7, 8, 9). This difficulty was not related to the absence of pitch information since their pitch perception threshold did not differ from other patient groups. It has been proposed that the right IFG plays a general role in voice recognition and social communication at large, since direct connections have been demonstrated between the STG and the ventrolateral pre-frontal cortex of the macaque (equivalent of IFG in humans) along with vocalization responsive cells (Romanski, 2012). Voice deficits observed here following right IFG lesions, often in association with insula lesions (although the association was not significant), could then reflect a disconnection syndrome from the STS.

Results observed in left hemisphere patients were heterogeneous. The absence of one-to-one mapping between speech deficits and brain lesions is not completely surprising since many studies have found that phonological categorization depends critically on the left supramarginal gyrus, a region not

investigated in our study. Phonological deficits were also not always associated with aphasia (4 out of 16 patients), which concurs with the idea that sub-lexical speech perception impairments do not necessarily predict auditory comprehension deficits (Turkeltaub & Coslett, 2010). More than half of the left hemisphere patients (9 out of 16) were either impaired or unimpaired in both tasks, and in the aphasic group 4 out of 10 patients showed a double deficit and 5 others showed reduced performances, resulting in lower perceived distances between extreme stimuli (either male–female or/pa/–/ta/). One possible explanation is that aphasic patients simply did not understand the instructions. Because these same patients could, however, perform other tasks with more complex instructions (e.g., the tone discrimination task used to check their pitch perception – see method), we consider this pattern of results as supportive of the co-optation hypothesis. Indeed, if we accept that there is speech-selectivity in the left fronto-temporal cortex, then we have to conceive that this selectivity can be associated with the mechanisms that produce and perceive the sounds of speech (McGettigan & Scott, 2012). The dissociations observed in 3 left hemisphere patients, also reveals that gender categorization deficits can be observed following left hemisphere lesions, and therefore that gender categorization is processed bilaterally.

In conclusion, through the analysis of categorization performances of right fronto-temporal stroke patients, we showed that the right frontal cortex (likely the ventral part of the IFG) plays a major role in voice/gender information processing. In contrast, left fronto-temporal patients (aphasic or not) tend to show associated performances for both voice/gender perception and speech perception, although dissociation are also possible. Together, these results lend support to the hypothesis of bilateral processing of voice information with (i) an important role of the right frontal cortex in voice categorization and (ii) both common and dedicated mechanisms, in the left hemisphere, for talker and speech information processing.

Acknowledgment

This work was supported by SINAPSE (<http://sinapse.ac.uk/>) who funded A.B. Jones.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.cortex.2015.07.004>.

REFERENCES

- Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Starchenko, M. G., Tervaniemi, M., et al. (2006). Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. *Brain Research*, 1075(1), 142–150.
- Andics, A., Gacsi, M., Farago, T., Kis, A., & Miklosi, A. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology: CB*, 24(5), 574–578. <http://dx.doi.org/10.1016/j.cub.2014.01.058>.
- Bamiou, D. E., Musiek, F. E., Stevens, J., Cipolotti, L., Brown, M. M., & Luxon, L. M. (2006). Auditory temporal processing deficits in patients with insular stroke. *Neurology*, 67(4), 614–619.
- Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1476), 2091–2107. <http://dx.doi.org/10.1098/rstb.2006.1933>.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1), 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(20), 309–312.
- Bestelemyer, P. E. G., Belin, P., & Grosbras, M.-H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, 21(20), R838–R839.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brumfitt, S. M., & Sheeran, P. (1999). The development and validation of the Visual Analogue Self-Esteem Scale (VASES). *The British Journal of Clinical Psychology/the British Psychological Society*, 38(Pt 4).
- Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2012). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex*, 23(4), 958–966. <http://dx.doi.org/10.1093/cercor/bhs090>.
- Fay, M. P. (2010). Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics*, 11(2), 373–374. <http://dx.doi.org/10.1093/biostatistics/kxp050>.
- Fecteau, S., Armony, J. L., Joanete, Y., & Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *Journal of Neurophysiology*, 94(3), 2251–2254.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, 111(38), 13795–13798. <http://dx.doi.org/10.1073/pnas.1401383111>.
- Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: a neuropsychological analysis. *Neuropsychologia*, 48(4), 1104–1114. <http://dx.doi.org/10.1016/j.neuropsychologia.2009.12.011>.
- Hailstone, J. C., Crutch, S. J., & Warren, J. D. (2010). Voice recognition in dementia. *Behavioural Neurology*, 23(4), 163–164. <http://dx.doi.org/10.3233/BEN-2010-0284>.
- Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, 1(3), 242–249. <http://dx.doi.org/10.1093/scan/nsi027>.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). Presented at the ECV, Perception. What's new in Psychtoolbox-3? (Vol. 36). <http://dx.doi.org/10.1068/v070821>.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195–1197. <http://dx.doi.org/10.1126/science.3629235>.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of Acoustical Society of America*, 49(2), 467–477.
- Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 65, 497–516. <http://dx.doi.org/10.2307/1418032>.

- Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243, 489–494.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends in Cognitive Sciences*, 16(5), 269–276. <http://dx.doi.org/10.1016/j.tics.2012.04.006>.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker contingent process. *Psychological Science*, 5(1), 42–46. <http://dx.doi.org/10.1111/j.1467-9280.1994.tb00612.x>.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 19(2), 309–328.
- Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, 3. <http://dx.doi.org/10.3389/fpsyg.2012.00023>.
- Pernet, C. R., Belin, P., & Jones, A. (2014). Behavioral evidence of a dissociation between voice gender categorization and phoneme categorization using auditory morphed stimuli. *Frontiers in Psychology*, 4, 1018. <http://dx.doi.org/10.3389/fpsyg.2013.01018>.
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., Watson, R. H., Fleming, D., Crabbe, F., Valdes-Sosa, M., & Belin, P. (2015). The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, 119, 164–174. <http://dx.doi.org/10.1016/j.neuroimage.2015.06.050>.
- Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia*, 45(8), 1899–1910. <http://dx.doi.org/10.1016/j.neuropsychologia.2006.11.015>.
- Perreault, C., & Mathew, S. (2012). Dating the origin of language using phonemic diversity. *PLoS One*, 7(4), e35289. <http://dx.doi.org/10.1371/journal.pone.0035289>.
- Rama, P., Poremba, A., Sala, J. B., Yee, L., Malloy, M., Mishkin, M., et al. (2004). Dissociable functional cortical topographies for working memory maintenance of voice identity and location. *Cerebral Cortex (New York, N.Y.: 1991)*, 14(7), 768–780. <http://dx.doi.org/10.1093/cercor/bhh037>.
- Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology. Human Perception and Performance*, 23(3), 651–666.
- Romanski, L. M. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proceedings of the National Academy of Sciences*, 109(Suppl. 1), 10717–10724. <http://dx.doi.org/10.1073/pnas.1204335109>.
- Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2014). Voice identity recognition: functional division of the right STS and its behavioral relevance. *Journal of Cognitive Neuroscience*. http://dx.doi.org/10.1162/jocn_a_00707.
- Scott, S. K. (2008). Voice processing in monkey and human brains. *Trends in Cognitive Sciences*, 12(9), 323–325.
- Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). *Journal of Speech and Hearing Disorders*, 45(3), 308–324.
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, 114, 1–15.
- Turner-Stokes, L., Kalmus, M., Hirani, D., & Clegg, F. (2005). The Depression Intensity Scale Circles (DISCs): a first evaluation of a simple assessment tool for depression in the context of brain injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76(9), 1273–1278. <http://dx.doi.org/10.1136/jnnp.2004.050096>.
- Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, 1(2), 185–195.
- Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 24(2).
- Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, 11(5), 665–674. <http://dx.doi.org/10.1080/01688638908400923>.
- 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 Research. Cognitive Brain Research*, 17(1).
- Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of Acoustical Society of America*, 118, 2618–2633. <http://dx.doi.org/10.1121/1.2011156>.
- Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., et al. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *Journal of Neuroscience*, 29(1), 61–69. <http://dx.doi.org/10.1523/JNEUROSCI.3489-08.2009>.
- Whiteley, W., Lindsey, R., Wardlaw, J., & Sandercock, P. (2006). Third International stroke trial. *International Journal of Stroke*, 1, 172–176.
- Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Oxford, UK: Academic Press, Elsevier.