



HAL
open science

Sound-Induced Activity in Voice-Sensitive Cortex Predicts Voice Memory Ability

Rebecca Watson, Marianne Latinus, Patricia E G Bestelmeyer, Frances
Crabbe, Pascal C Belin

► **To cite this version:**

Rebecca Watson, Marianne Latinus, Patricia E G Bestelmeyer, Frances Crabbe, Pascal C Belin. Sound-Induced Activity in Voice-Sensitive Cortex Predicts Voice Memory Ability. *Frontiers in Psychology*, 2012, 3, pp.89. 10.3389/fpsyg.2012.00089 . hal-02006948

HAL Id: hal-02006948

<https://hal.science/hal-02006948v1>

Submitted on 5 Feb 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.



Sound-induced activity in voice-sensitive cortex predicts voice memory ability

Rebecca Watson^{1*}, Marianne Latinus¹, Patricia E. G. Bestelmeyer², Frances Crabbe¹ and Pascal Belin^{1,3}

¹ Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

² School of Psychology, Bangor University, Bangor, UK

³ International Laboratories for Brain, Music and Sound, Université de Montréal and McGill University, Montreal, QC, Canada

Edited by:

Claude Alain, Rotman Research Institute, Canada

Reviewed by:

Erich Schröger, University of Leipzig, Germany

Kristina Simonyan, Mount Sinai School of Medicine, USA

*Correspondence:

Rebecca Watson, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK.
e-mail: r.watson@psy.gla.ac.uk

The “temporal voice areas” (TVAs; Belin et al., 2000) of the human brain show greater neuronal activity in response to human voices than to other categories of non-vocal sounds. However, a direct link between TVA activity and voice perception behavior has not yet been established. Here we show that a functional magnetic resonance imaging measure of activity in the TVAs predicts individual performance at a separately administered voice memory test. This relation holds when general sound memory ability is taken into account. These findings provide the first evidence that the TVAs are specifically involved in voice cognition.

Keywords: temporal voice areas, memory, identity, paralinguistic processing

INTRODUCTION

The human voice is probably the most common and meaningful sound of our environment, and efficient processing of the information carried within it – both linguistic and non-linguistic – is important for normal social interaction. However, in contrast to speech perception, relatively little is known about the neural bases of our ability to interpret different kinds of paralinguistic information from the voice – such as a speaker’s age, gender, or emotional state.

Functional magnetic resonance imaging (fMRI) studies have identified regions along the middle and anterior part of the superior temporal sulcus (STS) with a preferential response to vocal sounds (the “temporal voice areas,” TVA; Belin et al., 2000). Their early development (Grossman et al., 2010), ancient phylogenetic history (Petkov et al., 2008), and crucially, preferential response to vocalizations even devoid of linguistic content (Binder et al., 2000; Belin et al., 2002) suggest that the TVA could constitute a critical node of the cerebral network involved in voice cognition abilities. However, little is known of the exact functional role of the TVA, and whether their greater response to voice is indicative of a specific role in cerebral voice processing.

In the current study, our aim was to examine the relation between an aspect of vocal paralinguistic processing ability – memory for different speakers’ voices – and an fMRI measure of activity in the TVA. Deficits in voice recognition are a hallmark of phonagnosia, a condition in which an individual is unable to recognize familiar voices and shows impaired learning and recognition of new voices (Van Lancker et al., 1988; Garrido et al., 2009). However, other voice-related and auditory skills such as recognition of vocal emotions and processing of environmental sounds and music appear to remain intact, suggesting that the recognition of a speaker’s vocal identity depends on separable mechanisms from those used to recognize other information from the voice or non-vocal auditory stimuli. Although voice identification and

its neural correlates have received little scientific attention, there has been evidence proposing that auditory regions such as the STS play an important role in the representation of individual voices (Belin and Zatorre, 2003; Von Kriegstein et al., 2003). However, the link between voice memory ability – an important aspect of vocal identity processing – and voice-selective activity remains untested.

Rather than scan participants whilst performing a memory task, we chose to relate an offline measure of voice memory ability to passive TVA activity. The main focus of our experiment was not necessarily to examine the mechanisms of voice memory *per se*, but rather to test whether simply the strength of TVA activity in response to auditory stimulation, measured in an unrelated scan, could be in some way generally indicative of an individual’s ability to remember voices: a higher-level, cognitive skill. To further test the specificity of such a link between voice memory and TVA activity, we included a non-vocal control task with similar instructions (a bell-memory task). We asked whether a relation between TVA activity and voice memory performance could still be observed after accounting for variance associated with general sound memory ability.

MATERIALS AND METHODS

PARTICIPANTS

Fifty-two English-speaking adults (30 women; age \pm SD: 24 \pm 6 years), with self-reported normal audition, took part in one fMRI scan, and a separate behavioral assessment of vocal recognition. The behavioral test and the fMRI scan were not run simultaneously as our aim was not to directly investigate the cerebral correlates of voice memory encoding and retrieval *per se*, but rather to examine the link between activity in the voice-sensitive regions and performance on the voice memory test. The time interval between the behavioral test and the fMRI session varied between 0 (same day) and 490 days (mean \pm SD: 79 \pm 120 days).

Sixteen participants were run in the behavioral and fMRI experiment on the same day. Out of those who completed the test on different days, 8 were run firstly in the behavioral test (time interval between behavioral and fMRI scan: mean \pm SD: 18 \pm 17 days) and 28 firstly in the fMRI scan (time interval between fMRI scan and behavioral: mean \pm SD: 141 \pm 136 days). The ethical committee from the University of Glasgow approved the study, and all volunteers provided written informed consent before participating. Participants were paid a total of £15 to take part in both experiments.

BEHAVIORAL TESTING

The vocal memory test was designed as follows: (a) *Learning Phase*: participants heard eight different voices (four males and four females), each repeated consecutively three times, and were instructed to listen to these voices carefully in order to memorize them; (b) *Recognition Phase*: immediately after, participants listened to a second series of 16 voices (8 previously heard, 8 novel voices), and indicated after each one whether they had or had not heard it during the learning phase. Phases (a) and (b) were then repeated using bell sounds, in order to provide a measure of sound memory outside the vocal domain with similar setup and instructions.

Voice stimuli were obtained from previous recordings made in Montreal. The native language of all speakers was Canadian French. Eight male and 8 female speakers pronounced the French vowel “a” (“ah”), resulting in 16 different vocal stimuli for use in the memory test. It should be noted that our samples came from French speakers, whilst our participants were English speakers. It is possible that this difference in spoken language used could, for some participants, have caused impairment in performance; however, linguistic content of our samples was kept to a minimum, and the majority of our participants scored well above chance level. Recordings (16 bit) of the speakers were made in the multi-channel recording studio of Secteur ElectroAcoustique in the Faculté de musique, Université de Montréal, using two Brüel & Kjær 4006 microphones (Brüel & Kjær; Nærum, Denmark), a Digidesign 888/24 analog/digital converter and the Pro Tools 6.4 recording software (both Avid Technology; Tewksbury, MA, USA). The bell sounds were obtained from an internet source¹. They provided another set of sounds from a homogeneous category, with an effort to matching the sounds of the vocal set for variety.

Participants sat at a computer in a soundproof booth for the duration of the experiment. Sound stimuli were controlled using media control function (MCF; DigiVox, Montreal) via a computer in a separate room, and delivered through Beyerdynamic headphones at a comfortable sound-level.

fMRI SCANNING

Participants were run in a 10-min “voice localizer”². The voice localizer involves participants listening passively to 8-s blocks from either vocal or non-vocal sound categories presented with a 33% proportion of silent blocks in an efficiency-optimized

pseudo-random order. Vocal sounds were either speech (for example, isolated words, connected speech in several languages) or non-speech (such as laughs, sighs, and coughs) produced by several speakers of different gender and age. Non-vocal sounds included natural sounds, animal cries, mechanical sounds, instrumental sounds (including a number of bell sounds, all different from those of the memory test). The response to vocal as compared to non-vocal sounds is contrasted, in order to localize the TVAs.

It should be noted that the voice localizer scan is standard procedure for all participating in our laboratory’s experiments, and this data has been used in other published studies in order to localize the TVA (Latinus et al., 2011; Bestelmeyer et al., 2011).

Stimulus presentation was controlled with MCF (DigiVox, Canada). Stimuli were delivered binaurally through MRI-compatible electrostatic headphones (Nordic NeuroLab, Norway), at sound pressure level of approx. 80 dB SPL. Scanner noise was continuous throughout the experiment, providing a consistent auditory background.

Scanning was performed using a 3-T Siemens MRI system (Magnetom Vision, Siemens Electric, Erlangen, Germany) and an echoplanar gradient-echo pulse sequence (voxel size: 3 mm \times 3 mm \times 3 mm; TE = 30 ms; TR = 2000 ms; matrix = 64 \times 64). A total of 310 volumes per participant were acquired per session, with 32 slices per volume. High resolution, T1-weighted, anatomical images were also acquired for each participant (voxel size: 1 mm \times 1 mm \times 1 mm; TE: 2.52 ms; TR: 1900 ms; matrix size: 256 \times 256).

fMRI DATA PROCESSING AND ANALYSIS

Image processing and statistical analysis were carried out using SPM8³ (Statistical Parametric Mapping, Wellcome Department of Cognitive Neurology, London, UK). BOLD signal images were realigned, co-registered, normalized to the MNI template using a trilinear interpolation transformation and a voxel size of 3 mm \times 3 mm \times 3 mm, and then smoothed using an 8-mm FWHM Gaussian kernel. Condition related changes in regional brain activity were estimated for each participant by a general linear model in which the responses evoked by each condition of interest were modeled by a standard hemodynamic response function. The contrasts of interest were computed at the individual level to identify the cerebral regions significantly activated by all sounds (vocal + non-vocal sounds) relative to the periods of no stimulation (other than the scanner noise) used as a baseline. The contrast (vocal > non-vocal sounds) was computed to localize the TVAs. Cerebral activations were then examined at the group-level in random-effect analyses using one-sample *t*-tests. A group-level map of voice-preference was used as an indicator of TVA location in our group of participants and further used as an explicit mask (thresholded at $p < 0.05$; FWE) in further analyses.

In a first stage (Analysis 1), we identified regions within the TVA in which BOLD signal would be related to voice memory score. Percentages of correct answers for the voice memory task were related to: (a) BOLD signal representing the general sound-activation (vocal + non-vocal > baseline); (b) BOLD signal representing voice-preference (vocal > non-vocal).

¹www.findsounds.com

²http://vnl.psy.gla.ac.uk/resources/

³http://www.fil.ion.ucl.ac.uk/spm

In a second stage (Analysis 2), we asked whether the cluster identified in the first stage was specifically related to memory for voices. Due to the voice memory scores for bells and voices being highly correlated, we accounted for the residual variance associated *only* with the vocal memory task by orthogonalizing the scores for the vocal task, relative to the scores for the memory of bells. We then examined the relation between activity in the clusters identified in the first stage and a regressor which consisted of the orthogonalized voice memory scores.

In Analysis 1 our results were conducted at a threshold of $p < 0.001$ (uncorrected). Additionally, our analysis was restricted to the TVA – as opposed to the whole brain – due to the fact our intention was to investigate the link between voice perception behavior and voice-selective activity only. In Analysis 2 we used a small volume correction, with results thresholded at $p < 0.05$ (FWE). For both analyses we used a cluster size of five voxels or greater.

RESULTS

BEHAVIORAL RESULTS

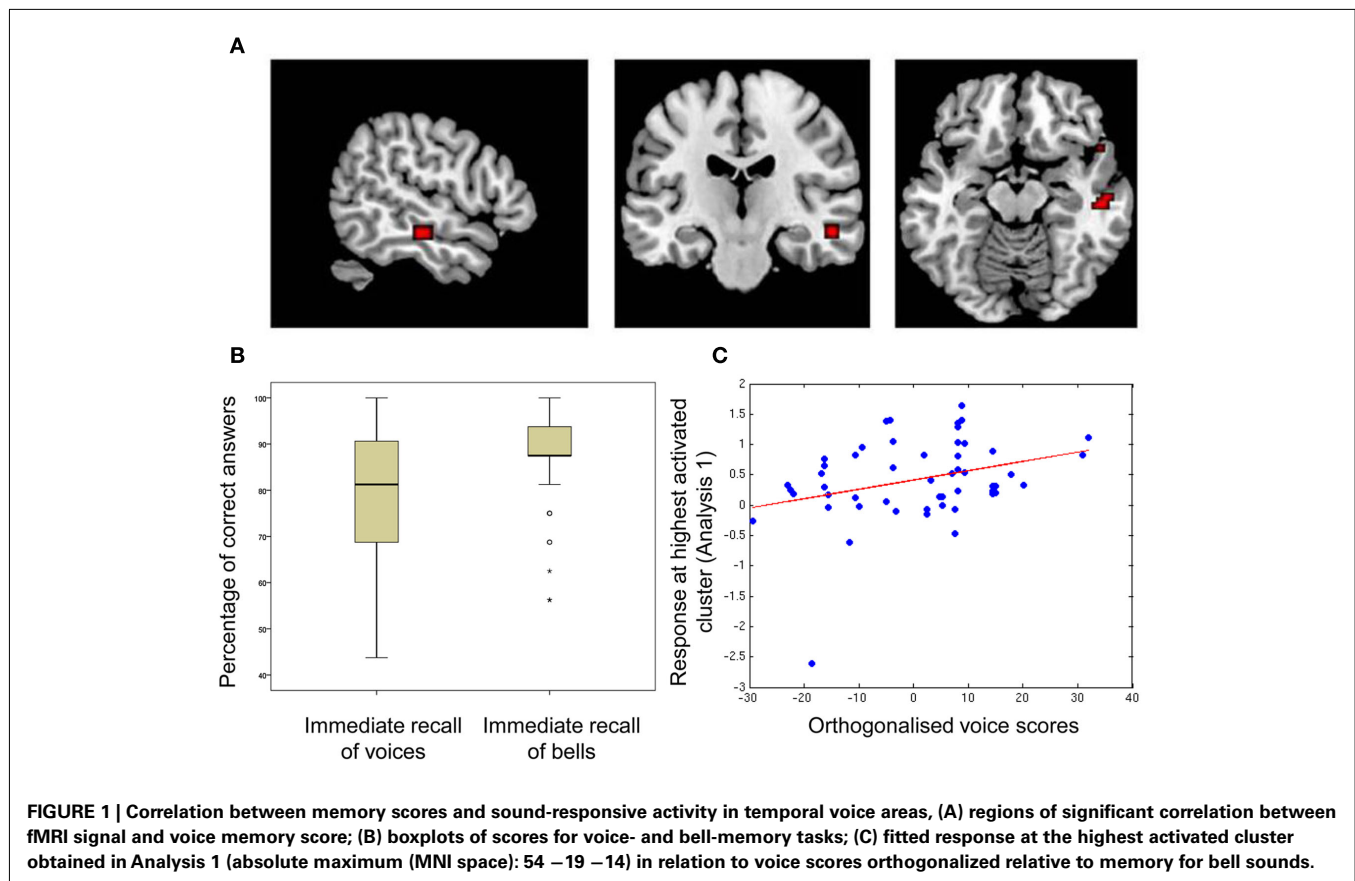
Participants' performance was significantly better for bell recognition [percent correct (mean \pm SD): 87.15 ± 11.1] than for voice recognition (78.54 ± 14.9 ; voice recognition vs. bell recognition; $t = -4.38$, $p < 0.0001$) but their performance at the two tasks was related [$r(50) = 0.40$, $p < 0.005$]. Boxplots of scores are shown in **Figure 1B**.

fMRI RESULTS

The contrast of BOLD signal measured during listening to the vocal vs. the non-vocal blocks of the voice localizer reliably localized the TVAs along the middle/anterior parts of the STS/STG bilaterally. Within the TVAs, relating voice memory score and general sound-activation (vocal + non-vocal > baseline; Analysis 1a) highlighted a region within the STS/MTG in which activation was correlated with voice memory performance (global maximum [$54 -19 -14$], $p < 10^{-4}$ (uncorrected), $t = 3.74$; cluster extending to 13 contiguous voxels; **Figure 1A**). However, relating voice memory score to voice-selectivity in the TVAs (vocal > non-vocal; Analysis 1b) did not highlight any significantly activated voxels [$p > 10^{-4}$ (uncorrected)].

Within the TVA cluster identified above, we observed a significant correlation between voice memory scores orthogonalized relative to the bells memory scores and BOLD activity [$p < 0.005$ (FWE), $t = 3.39$].

Within our group of participants, there was one outlier with regards to auditory response within the voice-selective regions (see **Figure 1C**). This participant was also one of the poorest performers in the immediate vocal recall task. Although their behavioral score was contained within the normal distribution, we were concerned that their markedly lower response within the TVA could be driving the effect we observed. Thus, we tested to see whether there was still a significant correlation after removal of this outlying participant. We related orthogonalized voice scores (not including



the outlier) to activity within the cluster identified in Analysis 1, and found a significant correlation still emerged [$p < 0.03$ (FWE), $t = 2.61$].

DISCUSSION

We find that performance at a voice memory test is well predicted by a separate measure of TVA activity obtained during passive listening to sounds: subjects with higher sound-induced activity in the TVA correctly remembered more voices, even when variance related to general sound memory performance had been included in the model. This constitutes the first evidence of a link between TVA activity and voice perception behavior, suggesting a critical role of the TVA in voice cognition.

Correlations between voice perception performance and brain activity in the vicinity of the TVA have already been reported in previous studies. For example, cerebral activity measured with positron emission tomography (PET) in the left frontal pole and right temporal pole has been found to correlate with participants' correct identification of familiar voices during the scan (Nakamura et al., 2001). However, it is not clear whether these activations are voice-specific or whether they would also be observed with other sound categories. Moreover, performance-dependent activations during a task could

be related to factors not specifically involved in processing that sound category (e.g., general arousal, attentional factors): the non-independent, simultaneous measurement of activity, and task performance in these studies means we cannot generalize the results to other tasks and situations. In contrast, that our behavioral results were obtained outside the scanner makes the observed correlation more compelling. Our predictive fMRI measure was obtained independently from the task being evaluated, during passive listening conditions, often on a separate day than the memory test; yet, it provides important information regarding at least one aspect of a person's "higher-level" voice perception abilities.

A notable finding is that the TVA's voice-preference (the degree to which they responded more to the vocal than to the non-vocal sounds) did not act as a predictor of voice memory performance. Rather, it was the general sound-induced activity within the TVA (measured using both vocal and non-vocal sounds) that was significantly related to voice memory performance. This observation suggests that it is the general synaptic density in the TVA (most strongly correlated with BOLD signal in response to sounds; Logothetis et al., 2001) rather than the proportion of voice-preferring neurons in the TVA (correlated with its voice-preference) that is most predictive of voice perception behavior.

REFERENCES

- Belin, P., and Zatorre, R. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14, 2105–2109.
- Belin, P., Zatorre, R. J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cogn. Brain Res.* 13, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Bestelmeyer, P. E. G., Latinus, M., Bruckert, L., Rouger, J., Carbbe, F., and Belin, P. (2011). Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. *Cereb. Cortex*. doi: 10.1093/cercor/bhr204
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.
- Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., Schweinberger, S. R., Warren, J., and Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia* 47, 123–131.
- Grossman, T., Oberecker, R., Kock, S. P., and Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron* 66, 1825–1828.
- Latinus, M., Crabbe, F., and Belin, P. (2011). Learning-induced changes in cerebral processing of vocal identity. *Cereb. Cortex* 21, 2820–2828.
- Logothetis, N., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., and Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* 39, 1047–1054.
- Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374.
- Van Lancker, D. R., Cummings, J. L., Kreiman, J., and Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex* 24, 195–209.
- Von Kriegstein, K., Eger, E., Kleinschmidt, A., and Giraud, A. L. (2003). Modulation of neural responses to speech by direction attention to voices or verbal content. *Cogn. Brain Res.* 17, 48–55.
- commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 December 2011; accepted: 08 March 2012; published online: 02 April 2012.

Citation: Watson R, Latinus M, Bestelmeyer PEG, Crabbe F and Belin P (2012) Sound-induced activity in voice-sensitive cortex predicts voice memory ability. *Front. Psychology* 3:89. doi: 10.3389/fpsyg.2012.00089

This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Watson, Latinus, Bestelmeyer, Crabbe and Belin. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any