CORTEX 192 (2025) 35–52

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Are the ventral anterior temporal lobes involved in accessing conceptual knowledge during spoken word production? fMRI evidence from auditory naming

Angelique Volfart [a,*,1], Katie L. McMahon [b,c], Catherine Liégeois-Chauvel [d], Vitória Piai [e] and Greig de Zubicaray [a]

[a] School of Psychology and Counselling, Faculty of Health, Queensland University of Technology, Australia
[b] School of Clinical Sciences, Faculty of Health, Queensland University of Technology, Australia
[c] Herston Imaging Research Facility, Royal Brisbane & Women's Hospital, Australia
[d] Cortical Systems Laboratory, Department of Neurological Surgery, School of Medicine, University of Pittsburgh, PA, USA
[e] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Article history:
Received 11 November 2024
Accepted 4 August 2025
Action editor Pia Rotshtein
Revised date: 30 July 2025
Published online 4 September 2025

Keywords:
Functional MRI
Multiband echoplanar imaging
Naming to definitions
Naming to environmental sounds
Spoken word production

ABSTRACT

There is ongoing debate about the role of ventral anterior temporal lobe (vATL) regions in the initial stages of production, particularly in accessing conceptual knowledge, with most evidence coming from visual naming tasks. Here, we investigated whether these regions are engaged during naming from different types of auditory stimuli. Twenty-five participants completed an fMRI experiment involving two naming tasks: auditory sentence definition (e.g., *The yellow part of an egg*) and nonverbal environmental sound (e.g., a sheep *bleating*). Our overall aim was to identify brain regions that are commonly activated across both naming tasks as well as those showing task-specific activations. With regards to the vATL's role, we hypothesised that these regions would show common activation across naming tasks, consistent with their proposed role in crossmodal conceptual processing, one of the first processing stages for retrieving a word based on an external input. Left-lateralized activation common to both tasks was observed in posterior fusiform gyrus, superior temporal gyrus and sulcus, inferior and superior frontal gyrus, and subcortical and cerebellar regions. Significant activation was observed in the bilateral vATLs only during naming to definitions, despite tSNR being equivalent across tasks. Our findings indicate that environmental sounds do not activate the vATL to the same extent as

* *Corresponding author*. A Wing, O Block, Kelvin Grove Campus, School of Psychology and Counselling, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia.
E-mail address: angelique.volfart@uni.lu (A. Volfart).
[1] Note the author has moved to University of Luxembourg since the work described in the article was carried out. Present address is: Department of Behavioural and Cognitive Sciences, Institute of Cognitive Science & Assessment, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

https://doi.org/10.1016/j.cortex.2025.08.010

auditory definitions, placing constraints on the crossmodal nature of semantic representations in these regions.

1. Introduction

Spoken word production is a core human skill, usually executed daily with great efficiency. There is general consensus among psycholinguistic theories that it involves several stages from selecting the intended meaning to express ("preverbal" conceptual preparation), retrieval of its lexical and form properties and, finally, articulation (Dell et al., 1997; Levelt, 1999; Levelt et al., 1999). The neural correlates of the multiple processes subtending word production are distributed primarily over the left perisylvian cortex. Reviews and meta-analyses have consistently implicated the middle and superior temporal gyri (MTG and STG, respectively), the inferior frontal gyrus (IFG), the precentral gyrus and, more recently, the inferior parietal lobule (IPL) as important regions for mediating the processing stages that occur *after* conceptual preparation (e.g., Indefrey, 2011; Price, 2012; Kemmerer, 2019; Nozari, 2022; de Zubicaray, 2023 for reviews). Yet, despite many neuroimaging and lesion studies concluding that ventral anterior temporal lobe (vATL) regions, including the anterior sections of the inferior temporal and fusiform gyri (ITG and FG, respectively; with a locus usually centred around MNI coordinates ±36, −15, −30, Binney et al., 2010; see also Rice et al., 2018), are critical for representing conceptual knowledge (e.g., Patterson et al., 2007; Binney et al., 2010; Mion et al., 2010; Visser et al., 2012; see also Lambon Ralph et al., 2017), these regions are usually only briefly mentioned in neurobiological accounts of spoken word production and their precise role in processing conceptual representations remains debated.

To date, most of the neuroimaging, lesion and brain stimulation evidence supporting the neurobiology of conceptually driven spoken word production has been derived from naming paradigms. The overwhelming majority of these studies have employed visual naming, followed by auditory naming from either verbal descriptions/definitions (e.g., *The thing that unlocks a door*) or environmental sounds (e.g., a cat *meowing*). The "lead−in" (e.g., Indefrey & Levelt, 2004) processes differ across paradigms, with visual naming relying on visuospatial and object recognition processes and auditory naming relying on linguistic speech or nonverbal sound recognition.

According to production accounts, modality-specific lead−in processes activate preverbal concepts which in turn spread their activation to lexical representations via conceptual-to-lexical connections (Dell et al., 1997; Levelt, 1999). However, unlike pictures and environmental sounds that access their meanings directly from their respective visual and auditory percepts, spoken words access their meanings indirectly via lexical representations, a process which may or may not involve simulation (Barsalou, 2020; Brownsett et al., 2022; Chen & Spence, 2018; Glaser & Glaser, 1989). Naming response times also differ across stimuli, with pictures being responded to more quickly than environmental sounds (e.g., Brownsett et al., 2022; Tranel et al., 2005) and spoken definitions (e.g., Hamberger et al., 2021; Hamberger & Seidel, 2003). In addition, environmental sounds are viewed as having weaker conceptual-to-lexical connections than depictions of object category exemplars (Brownsett et al., 2022; Wöhner et al., 2020) and so might be less strongly activated, explaining their relatively slower responses during production. Despite the differences across the three naming paradigms, processing across all is assumed to converge on shared, preverbal semantic representations. The nature of these latter representations remains a matter of debate, although recent reviews (de Zubicaray, 2023; Kemmerer, 2019; Nozari, 2022) have tended to adopt proposals for cross-modal (amodal or heteromodal) semantic representations, informed by neurobiological accounts emphasising a role for the vATL in particular as a semantic convergence zone or "hub" (e.g., Lambon Ralph et al., 2017).

While a role for the vATL in semantic processing has been consistently demonstrated in comprehension tasks, evidence for this region's involvement during production is sparse and mostly limited to visual naming tasks (e.g., Hoffman and Lambon Ralph 2018). In addition, several studies have shown that while the brain networks for language comprehension and production largely overlap, some areas are preferentially engaged depending on the task (see, e.g., Price, 2012; Walenski et al., 2019; but see also Fairs et al., 2022 for a recent discussion on this topic). For example, in her meta-analysis, Price (2012) noted that the posterior ITG was primarily associated with word retrieval tasks. More recently, a study investigating activation overlap between discourse speaking and listening tasks found a correlation between activation in the bilateral vATL and semantic content only for comprehension (Patel et al., 2023).

These limited findings regarding the vATL's role in production indicate that further investigation is needed beyond visual naming. Below we summarise the available evidence concerning the involvement of ventral temporal lobe regions in "multimodal" conceptually driven spoken word production from studies comparing naming tasks from different input types within participants.

1.1. *Evidence from clinical populations*

Most neurobiological investigations comparing visual and auditory naming with verbal definitions or nonverbal environmental sounds have been conducted in people with intractable epilepsy using intracranial electroencephalography (iEEG) recordings and direct electrical stimulation (DES).

Comparisons of visual naming and naming to definition in these studies have yielded somewhat inconsistent results, showing common involvement of either anterior or posterior ventral temporal lobe regions (e.g., Cervenka et al., 2013; Forseth et al., 2018; Malow et al., 1996; Nakai et al., 2019) or no common engagement (e.g., left anterior ITG sites disrupting definition naming only in Hamberger et al., 2001; high-gamma modulations in the bilateral FG only for visual naming in Kojima et al., 2013; Ikegaya et al., 2019; Nakai et al., 2019; Kitazawa et al., 2023). To our knowledge, only two studies have administered the three different naming tasks (naming to environmental sounds, verbal definitions and pictures) in the same group of patients. An iEEG study of high−gamma activity observed no common sites, with activity observed in the bilateral FG and ITG only for the latter task (Kitazawa et al., 2023), while a study of patients with the semantic variant of primary progressive aphasia (semantic dementia) reported common correlations with grey matter volume of the left FG (Ding et al., 2016).

In summary, clinical studies provide only equivocal evidence for common ventral temporal lobe engagement across visual and auditory naming paradigms, with the limited number of positive findings showing considerable variability over the anterior-to-posterior axis.

1.2. *Evidence from healthy participants*

Only a few neuroimaging studies have compared visual and auditory (verbal definition) naming tasks in the same group of participants, with mixed results. An fMRI study using covert production reported a common area of activation for both tasks only in the left posterior temporal lobe (no coordinate provided) (Hamberger et al., 2014). Forseth et al. (2018) also administered picture and definition covert naming tasks to a group of healthy participants during fMRI and observed similar results to their sample of epileptic patients, i.e., common activation in the left middle fusiform gyrus (around MNI $y = -17$, similar to the *y*-axis coordinate reported by Binney et al., 2010, although located more medially on the *x*-axis).

The neural correlates of naming to environmental sounds have rarely been explored in healthy participants. To the best of our knowledge, only two studies have compared sounds with pictures in this population. Using positron emission tomography (PET), Tranel et al. (2005) investigated overt naming of animals and tools from visual or auditory inputs and reported common activation for naming animals from left ventral anterior to posterior temporal cortex (Talairach = −27, −9, −29[2] and Talairach = −36, −37, −16,[3] respectively) although common activation for naming of tools was observed only in the left ventral posterior temporal cortex (Talairach = −34, −36, −12[4]). This category-specific pattern of activation within the vATL appears inconsistent with the proposal of a unified semantic hub responsible for processing stimuli across categories and modalities (see Pobric et al., 2010). However, it is worth noting that the coordinates reported by Tranel et al. (2005) are situated more anteriorly than those typically associated with the ATL hub in the broader literature (e.g., Binney et al., 2010; Rice et al., 2018). Using fMRI, Reilly et al. (2016) reported that naming of familiar versus novel environmental sounds or pictures did result in significant activation of the lateral ATL (anterior MTG) but its ventral aspect was not specifically investigated. In addition, two other neuroimaging studies investigated environmental sound naming alone: one covert production fMRI study that did not report any activation in ventral temporal regions (Tomasino et al., 2015) and one PET study with overt production that did (Tranel et al., 2003, with a locus in the left posterior ITG: Talairach = −38, −28, −15[5]).

Overall, the limited evidence indicates posterior ventral temporal regions are relatively reliably engaged in cross-modal naming, consistent with a possible role in representation of conceptual knowledge. However, the same conclusion cannot be made for the vATL. However, none of the above-mentioned fMRI studies employed acquisition approaches to compensate for distortions and signal dropouts over the vATL, which is likely to have reduced their likelihood of finding activity in this region. Hence, a more targeted investigation of the vATL's role during auditory naming tasks is needed.

1.3. *Methodological challenges for fMRI investigations of spoken word production*

A major limitation of using fMRI to investigate the role of vATL regions in spoken word production is that these areas (and orbitofrontal cortex) are prone to magnetic susceptibility artefacts/distortions and partial signal dropout during conventional single-band gradient-echo (GE) echo planar imaging (EPI) fMRI acquisitions (e.g., Embleton et al., 2010; Ojemann et al., 1997). These problems are further exacerbated by speech-related motion artefacts (Mehta et al., 2006). This is perhaps why *all* of the previous fMRI studies of multimodal naming reviewed above employed covert rather than overt production. Yet, in their meta-analysis of neuroimaging studies of spoken word production conducted two decades ago, Indefrey and Levelt (2004) had already noted that ventral temporal regions were less frequently reported in fMRI and PET studies using covert compared to overt production.

In recent years, production researchers have increasingly adopted multiband EPI sequences for fMRI investigations employing overt naming (e.g., Roos et al., 2023; Todorović et al., 2023; Volfart et al., 2024), as they afford a number of advantages over single-band EPI acquisitions (for a review, see Wall, 2023). These include increased signal-to-noise (SNR) and temporal resolution that permit sampling of respiration related signal changes and pulsatile motion of cerebrospinal fluid (CSF) associated with cardiovascular cycles, reduced scanner acoustic noise, and improved statistical power due to the increased sampling rate (Chen et al., 2015; Demetriou et al., 2018; Feinberg et al., 2010). When used in conjunction with distortion-correction and denoising techniques, multiband EPI sequences have been shown to provide adequate

[2] To enable comparisons across studies, we converted original Talairach to MNI coordinates using the BioImage Suite Web (version 1.2.0; Papademetris et al., 2006). For this study, conversion to MNI gives a centroid at −28, −6, −38.

[3] Conversion to MNI coordinates: −39, −36, −23.

[4] Conversion to MNI coordinates: −36, −35, −19.

[5] Conversion to MNI coordinates: −42, −27, −22.

coverage of ventral temporal regions to identify activation during overt naming (e.g., Volfart et al., 2024).

1.4. The present study

To clarify the role of the vATL during spoken word production, we employed multiband fMRI to investigate healthy participants' performance on two auditory naming tasks: naming to definitions and nonverbal environmental sounds. Our overall aim was to identify brain regions that are commonly activated across both naming tasks as well as those showing task-specific activations attributable to the different lead–in processes (linguistic versus nonverbal). Following proposals that bilateral vATL regions play a role in representing preverbal conceptual knowledge during multimodal naming (e.g., Lambon Ralph et al., 2017), we hypothesized that these regions should show common activation during both auditory naming tasks consistent with this role.

2. Materials and methods

2.1. Participants

Twenty-eight healthy right-handed native English speakers were initially recruited, of whom three were later excluded from data analysis (one due to technical issues, two due to incidental findings). The final sample included 25 participants (13 females, mean age = 23.34 ± 5.2) who all reported normal or corrected-to-normal vision, and no hearing deficits, history of neurological or psychiatric disorder or learning disability or use of psychotropic medication (e.g., antidepressants). We did not conduct a formal power analysis to determine the sample size. Rather, we ensured our sample exceeded those of previous fMRI studies of auditory naming in healthy participants (13 participants, Hamberger et al., 2014; 18 participants, Reilly et al., 2016; 21 participants, Forseth et al., 2018). All inclusion and exclusion criteria were established prior to data analysis. All participants gave written consent before taking part in the experiment. After the study, participants received AUD$30 to reimburse their participation in the study. The study was approved by Metro North Health (HREC/15/QRBW/447) and QUT Human Research Ethics Committees (HE 2015-1148-1620).

2.2. Stimuli

For the naming-to-definitions task, stimuli were identical to those described in our recent study (Volfart et al., 2024). Fifty definitions were selected from three previous studies: 15 from Hamberger and Seidel (2003), 22 from Yochim et al. (2015) and 13 from Hamberger et al. (2021). Four additional definitions were selected to be used as practice stimuli.

All definitions were read aloud by a native Australian English female speaker while being digitally recorded with a sampling rate of 44.1Hz (32-bit, monophonic). Audio files containing definitions were normalized, cleaned and their volume equalized in Audacity (https://www.audacityteam.org/). Mean duration of resulting definition audio files was 2735 ± 583 ms (range: 1271–4083). Fifty control stimuli were created by reversing the audio track of each definition (i.e., playing it backward) to remove its meaning while preserving its acoustical properties. Two of the example definitions were also transformed with the same method to create control practice stimuli.

For the naming-to-environmental-sounds task, fifty stimuli were taken from two normative sound databases: 35 from Marcell et al. (2000) and 15 from Hocking et al. (2013). Four additional sounds were chosen to serve as practice. Mean duration of sound audio files was 2148 ± 1424 ms (range: 358–6084 ms). As audio files came from two different databases, we normalized and resampled them to 44.1 Hz in Audacity. In reversing the audio files as per the naming-to-definitions task, some recognizable features of the sound were preserved. Therefore, fifty control stimuli were created by generating white noise of constant amplitude based on the maximal amplitude in the original sound file and of the same duration as the original file.

Definition and environmental sound target words were matched on various psycholinguistic variables (word length, SUBTLEX word frequency, number of orthographic neighbours, number of phonological neighbours, concreteness, age of acquisition, number of phonemes, number of syllables) taken from the eLexicon database (Balota et al., 2007; https://elexicon.wustl.edu/index.html) (Table 1). In both tasks, the objects to be named were exemplars from a mix of living and non-living categories, following previous fMRI studies (e.g., Forseth et al., 2018; Hamberger et al., 2014; Reilly et al., 2016). As we noted above, the stimuli for both tasks were selected from previous work to ensure the use of well-validated stimuli. However, the semantic categories of the stimuli were not matched across tasks. This was due to practical constraints: many environmental sounds (e.g., musical instruments, nonverbal human sounds) did not have a verbal definition equivalent, and many definitions referred to concepts that were not associated with a specific or identifiable sound. As such, while all stimuli represented concrete concepts and required noun responses, the semantic categories differed between tasks. We will return to this issue in the Discussion section.

Given that most auditory stimuli that we selected were initially validated among American English speakers (except for Hocking et al. (2013)'s sound stimuli validated in Australian English speakers), we first conducted an online pilot study to ensure that these stimuli would elicit accurate responses in an independent sample of Australian English speakers (see

Table 1 – Psycholinguistic characteristics of target words to be produced in each auditory naming task.

	Definitions	Environmental sounds
Length (N letters)	6.1 (2.0)	6.42 (2.0)
Word frequency	10.49 (14.3)	12.19 (13.8)
N orthographic neighbours	2.2 (3.0)	2.16 (2.8)
N phonological neighbours	4.5 (6.9)	4.56 (6.3)
Concreteness ratings	4.76 (.3)	4.75 (.3)
Age of acquisition	6.18 (1.3)	5.67 (1.6)
N phonemes	5.18 (1.6)	5.24 (1.8)
N syllables	1.98 (.8)	1.96 (.9)

Supplementary Material). Results were used to define correct responses for the fMRI study.

2.3. Procedure

Before starting each task (naming to definitions or to environmental sounds), participants were instructed to carefully listen to each auditory stimulus and wait for a visual response cue to appear before naming, as quickly as possible, what they thought the definition or environmental sound corresponded to. They were also instructed to produce the word "noise" when hearing control stimuli (i.e., backward definitions or white noise).

After the completion of six practice trials, the main task then started, with each trial structured as follow: first, a black fixation cross was displayed in the center of a white background for 1500 ms and was followed by the stimulus presentation through headphones (duration ranging from 1271 to 4093 ms for definitions and from 358 to 6084 ms for environmental sounds). After a random jitter of 500–1500 ms during which a blank screen was presented, a visual response cue ("?") was displayed in the centre of the screen for 3000 ms and the microphone started recording the participant's verbal output. An inter-trial blank screen was then presented for 500 ms before starting the next trial (Fig. 1).

Each of the two auditory naming tasks comprised two blocks of 50 trials, with brief (< 3 min) rest breaks in-between, for a total of four blocks and 200 trials. Trials were pseudo-randomized in Mix (van Casteren & Davis, 2006) to ensure no more than two stimuli from the same condition (experimental or control) appeared consecutively. The task order was counterbalanced across subjects. The whole paradigm lasted about 30 min.

2.4. Apparatus

A PC running PsychoPy (https://www.psychopy.org/, version 2022.2.5) was used to present the auditory stimuli (via Optoacoustics Active II noise-cancelling headphones, Optoacoustics Ltd., Israel) and record the participant's verbal responses (via FOMRI III + noise-cancelling microphone, Optoacoustics Ltd., Israel) on digital audio files (48 kHz, monophonic). Visual cues were displayed on a white screen back projected to the participant with a BOLDScreen 32 LCD projector (Cambridge Research Systems Ltd., UK) and a mirror attached to the head coil.

2.5. Analysis of behavioural responses

Participants' responses to each item were transcribed from the digital audio files, and accuracy scored. Audio files were cleaned in Audacity to remove MRI noise and amplify the verbal output when necessary. Response times (RT) were automatically extracted via Chronset (https://www.bcbl.eu/databases/chronset; Roux et al., 2017) and further manually validated using a custom Praat script (van Scherpenberg et al., 2020, https://www.fon.hum.uva.nl/praat/).

For behavioral analyses, we only considered RT for correct responses and excluded outliers (3 SD from mean) for each participant and condition. Two-by-two repeated measures ANOVA were computed in SPSS (version 28.0.1.0, IBM Corp.) with factors Task (Definitions or Sounds) and Condition (Experiment or Control) to assess accuracy and RT performance across conditions.

2.6. MRI acquisition

A 3-T Siemens Magnetom Prisma MRI scanner (Siemens Healthineers, Erlangen, Germany) equipped with a 64-channel head-neck coil was used to acquire neuroimaging data at the Herston Imaging Research Facility. High-resolution 3D T1-weighted anatomical images were first collected using a MPRAGE sequence (TR = 1900 ms, TE = 2.32 ms, TI = 900 ms, flip angle (FA) = 9°, matrix = 256x256x192, FOV = 240x240 × 172.8 mm^3). Whole brain functional images were then acquired using a multiband single-echo EPI (twice refocussed) sequence with anterior-posterior (A-P) phase encoding and oblique slices optimized to obtain signal over anterior and ventral temporal lobe regions (Volfart et al., 2024; TR = 1000 ms, TE = 30 ms, FA = 80°, matrix = 72x72,

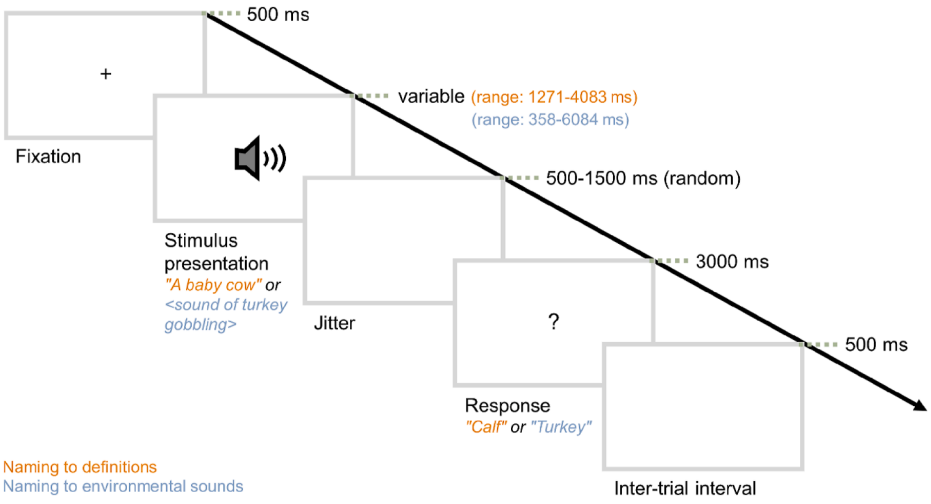

Fig. 1 – Schematic representation of experimental paradigm performed in the scanner.

FOV = 190 × 190mm^2, slices = 2.6 mm x 48, multiband acceleration factor = 4, bandwidth 2572 Hz/Px). The four functional runs (one per experimental block) lasted for 7:49 min for naming to definitions and 7:29 min for environmental sounds, with each including 10 s of dummy scans. A single set of field maps was acquired at the end of the session (TR = 520 ms, TE 1 = 4.92 ms, TE 2 = 7.38 ms, FA = 60°, matrix = 72x72, slices = 2.6 mm x 50, FOV = 190 × 190mm^2) to create a voxel displacement map (VDM). A T2-weighted image was also acquired after the functional runs for radiological review of incidental findings (TR = 6300 ms, TE = 100 ms, FA = 150°, matrix = 512x512, slices = 4.0 mm x 30, FOV = 220 × 220mm^2).

2.7. *Image analysis*

Data preprocessing and analyses were conducted using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) and the CONN toolbox version 22.a (https://web.conn-toolbox.org/; Nieto-Castañón & Whitfield-Gabrieli, 2022). Following the same procedure as described in Volfart et al. (2024), we first created a VDM in SPM12 using the FieldMap toolbox before preprocessing in CONN. Functional and anatomical data were then entered into a flexible preprocessing pipeline (Calhoun et al., 2017; Nieto-Castañón, 2022) beginning with the realign and unwarp function (Andersson et al., 2001) using the VDM to realign the timeseries to the mean image across the four functional runs. Potential outlier scans were identified using ART (Whitfield-Gabrieli et al., 2011) based upon framewise displacement above .9 mm or global BOLD signal changes above 5 standard deviations (Nieto-Castañón, 2022; Power et al., 2014). The mean functional image was then used to coregister the realigned and unwarped series to the T1-weighted image, which was next segmented into grey matter, white matter, and CSF tissue classes and normalized into standard MNI space using SPM12's unified segmentation and normalization algorithm (Ashburner, 2007; Ashburner & Friston, 2005) with the default IXI-549 tissue probability map template. The resulting transformation to MNI space was then used to normalise the functional data, after which they were resampled to 2 mm isotropic voxels and smoothed using spatial convolution with a Gaussian kernel of 8 mm full width half maximum (FWHM). In a final preprocessing step, we removed global signal changes arising from both intracranial and extracranial sources using a voxel-level linear model (LMGS; https://github.com/paulmmacey/lmgs/tree/master; Macey et al., 2004), with the residual time course data then entered into analysis. We performed this global signal regression (GSR) step as speech-related respiration changes manifest as significant decreases and increases in global signal intensity during and several seconds following the speech envelope, respectively, with extracranial signals observable in the vicinity of the temporalis muscle also correlated with global signal intensity changes during the speech envelope (Mehta et al., 2006). The application of GSR has been shown to result in dramatic improvements in temporal SNR in auditory naming (Volfart et al., 2024).

We conducted two-stage, fixed and random-effects model statistical analyses in SPM12. At the participant fixed-effects level, events corresponding to the start of the auditory stimulus presentation for correct trials in each task were modelled as effects of interest. The start of the response cue, the participants' response onsets (see section 2.5 above), and corresponding error trials were modelled as regressors of no interest. As recommended in Volfart et al. (2024), we also included the 6 rigid-body realignment parameters (RP), outlier scans (scrubbing parameter), and CSF timeseries (5 CompCor noise components) as regressors of no interest.

Our primary analysis aimed to identify common brain regions activated by both tasks (in comparison to their respective control stimuli). At the group random-effects level, we thus conducted a full factorial analysis using the Definition vs. Control and Environmental Sound vs. Control contrasts from each participant's fixed effect analysis and tested the conjunction (null) as defined by Nichols et al. (2005). Conjunction refers to the joint refutation of multiple null hypotheses (Friston et al., 2005). This allowed us to infer that there is a conjunction of all definition and environmental sound effects related to their processing (i.e., k = n; see Friston et al., 2005). Our secondary analysis aimed at identifying brain regions that were specifically activated in one task but not in the other. For this purpose, we exclusively masked each contrast of interest with the other contrast at $p < .001$ (uncorrected mask value).

We conducted unrestricted whole brain analyses as well as small volume correction (SVC; Worsley et al., 1996) targeting the bilateral vATL regions-of-interest (ROI) as identified previously in the context of semantic processing (left and right vATL; 6 mm radius spherical ROIs at MNI coordinates ±36, −14, −30[6] following the cluster identified by Binney et al., 2010; see also Rice et al., 2018, Fig. 3A). In addition, we performed an independent ROI analysis by extracting the mean beta values according to stimulus type from within each a priori defined spherical ROI based on the full factorial design using the MarsBar toolbox (Brett et al., 2002).

Clusters showing significant cortical BOLD signal changes in the whole brain analyses were labelled using the Brainnetome atlas parcellation scheme (https://atlas.brainnetome.org/brainnetome.html; Fan et al., 2016) and with the CoBrA atlas for clusters in the cerebellum (https://www.cobralab.ca/cerebellum-lobules; Park et al., 2014). We adopted a height threshold of $p < .001$ in conjunction with a spatial cluster extent threshold of $p < .05$ (family-wise error [FWE] corrected).

3. Results

3.1. *Behavioural results*

Behavioural results are presented in Table 2.

A 2x2 repeated measures ANOVA of accuracy revealed significant effects of Task [$F(1,24) = 37.292$, $p < .001$], Condition [$F(1,24) = 236.002$, $p < .001$], and their interaction [$F(1,24) = 73.325$, $p < .001$]. Accuracy was higher for naming to

[6] The original coordinates of the ROI reported by Binney et al., 2010 were ±36, −15, −30. Here the y coordinate was rounded to −14 due to the data having been resampled to 2 mm isotropic voxels. The results did not change when using y = −16 to define the ROIs.

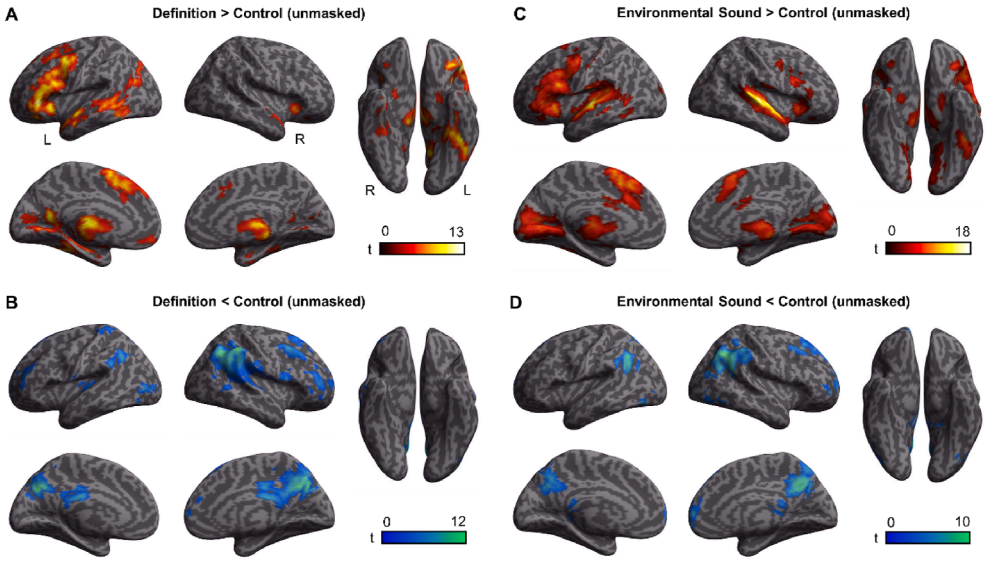

Fig. 2 – A. Significant BOLD signal changes in Definition > Control. B. Significant BOLD signal changes in Definition < Control. C. Significant BOLD signal changes in Environmental Sound > Control. D. Significant BOLD signal changes in Environmental Sound < Control. A-D results are shown on inflated surface renderings from SPM12. Height thresholded at $p < .001$ with a spatial cluster extent threshold at $p < .05$ (FWE-corrected).

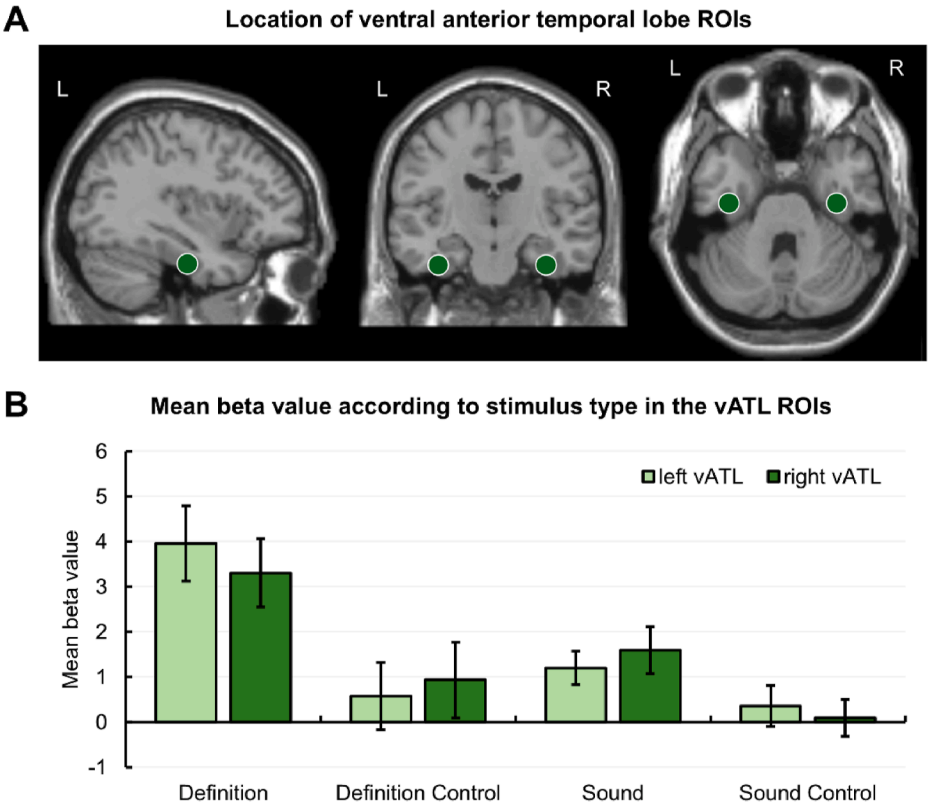

Fig. 3 – A. The ventral anterior temporal lobe regions-of-interest shown overlaid on the canonical single-subject MNI brain from SPM12. B. Plot showing the mean beta values from the left and right vATL ROIs according to stimulus type. Error bars represent standard errors.

Table 2 – Mean (standard deviations) of participants' correct response times (RT) in milliseconds and percent accuracy for naming to definitions and to environmental sounds.

	Naming to definitions		Naming to environmental sounds	
	Experimental items	Control items	Experimental items	Control items
Accuracy	89.6 (5.5)	99.6 (.8)	74.4 (9.4)	98.2 (5.2)
RT	602 (91)	688 (127)	644 (93)	666 (94)

definitions compared to environmental sounds and for control compared to experimental items.

The same analysis performed on RTs revealed a significant effect of Condition [$F(1,24) = 25.799$, $p < .001$] and interaction [$F(1,24) = 21.526$, $p < .001$]. The main effect of Task was not significant [$F(1,24) = .732$, $p = .401$]. Responses were quicker for experimental compared to control items overall and quicker for naming to definitions compared to environmental sounds.

3.2. Imaging results

3.2.1. Experimental compared to control stimuli

3.2.1.1. Unrestricted whole-brain analyses. The contrast Definition > Control revealed significant BOLD signal changes peaking in the left rostral inferior frontal gyrus (IFG, Brainnetome region labelled A45r, extending medially towards subcortical regions), the bilateral anterior superior temporal sulcus (STS), the left medial orbital gyrus (A14m), the right rostroventral fusiform gyrus (A20rv) and in three areas of the cerebellum (right cerebellar lobule VI and bilateral cerebellar white matter) (Fig. 2A and Table 3).

The contrast Definition < Control revealed significant BOLD signal changes peaking in the left dorsomedial parietooccipital sulcus (POS, in precuneus, extending towards the right POS and inferior parietal lobule, IPL, A40c), the left caudal IPL (A40c), the right inferior frontal junction (in middle frontal gyrus, MFG), the left area V5/MT (in lateral occipital cortex, LOC), the left MFG (A46), the left TE1.0 and TE1.2 area (in superior temporal gyrus, STG), the left postcentral superior parietal lobule (SPL, A7) and in the left cerebellar lobule Crus I (Fig. 2B and Table 3).

The contrast Environmental Sound > Control revealed significant BOLD signal changes peaking in the right caudal STG (A22c, extending towards the left STG), the left medial superior frontal gyrus (SFG, A8m) and right cerebellar lobule VI (Fig. 2C and Table 3).

The inverse contrast Environmental Sound < Control revealed significant BOLD signal changes peaking in the bilateral rostroventral IPL (A39rv), the right precuneus (A31), the right medial SFG (A10m) and the bilateral cerebellar lobule Crus I (Fig. 2D and Table 3).

3.2.1.2. SVC and independent ROI analyses. The contrast Definition > Control revealed significant clusters in the left (peak MNI = −38, −18, −26, cluster size = 57, z-score = 5.58) and right (peak MNI = 34, −10, −34, cluster size = 38, z-score = 4.27) vATL ROIs using SVC.

Table 3 – Brain regions showing significant BOLD increases and decreases for contrasts Experimental *vs.* Control.

	Peak MNI (x y z)			Z-score	Cluster size (N voxels)
Naming to definitions > control					
Left rostral A45 (IFG)	−48	30	0	Inf	21,544
Left anterior STS	−54	−6	−18	Inf	1240
Right cerebellar lobule VI	14	−74	−26	6.95	1075
Right cerebellar white matter	26	−40	−46	5.45	390
Left cerebellar white matter	−2	−60	−26	5.08	331
Left medial A14 (orbital gyrus)	−6	50	−14	5.00	175
Right rostroventral A20 (FG)	32	−8	−36	4.83	188
Right anterior STS	58	2	−12	4.65	314
Naming to definitions < control					
Left dorsomedial POS (precuneus)	−8	−68	32	Inf	10,911
Left caudal A40 (IPL)	−56	−58	38	6.47	803
Right inferior frontal junction (MFG)	44	12	48	6.12	3598
Left V5/MT (LOC)	−48	−82	2	5.64	438
Left cerebellar lobule Crus I	−46	−74	−38	5.60	370
Left A46 (MFG)	−38	58	10	5.36	836
Left TE1.0 and TE1.2 (STG)	−42	−22	8	4.98	490
Left postcentral A7 (SPL)	−26	−42	64	4.58	314
Naming to environmental sounds > control					
Right caudal A22 (STG)	64	−20	6	Inf	30,332
Left medial A8 (SFG)	−2	10	56	Inf	3494
Right cerebellar lobule VI	32	−60	−28	6.90	897
Naming to environmental sounds < control					
Right rostroventral A39 (IPL)	48	−60	38	7.52	3817
Right A31 (precuneus)	10	−50	32	6.90	3687
Left rostroventral A39 (IPL)	−50	−58	34	6.58	1036
Left cerebellar lobule Crus I	−42	−78	−34	6.22	658
Right medial A10 (SFG)	2	60	4	5.87	3360
Right cerebellar lobule Crus I	30	−86	−30	5.49	185

Note: Whole-brain analysis, height threshold at $p < .001$ and spatial cluster extent threshold at FWE $p < .05$ (i.e., 175 voxels for Definition > Control; 314 for Definition < Control; 897 for Environmental Sounds > Control; 185 for Environmental Sounds < Control). Peak locations were determined using the Brainnetome atlas, except for peak clusters in the cerebellum that were located using the CoBrA atlas (Park et al., 2014). FG = fusiform gyrus; IFG = inferior frontal gyrus; IPL = inferior parietal lobule; LOC = lateral occipital cortex; MFG = middle frontal gyrus; POS = parietooccipital sulcus; SFG = superior frontal gyrus; SPL = superior parietal lobule; STG = superior temporal gyrus; STS = superior temporal sulcus.

No significant clusters were detected in the vATL ROIs for the contrasts Definition < Control, Environmental Sound > Control and Environmental Sound < Control using SVC.

Fig. 3B shows the mean beta values according to stimulus type extracted from within each spherical ROI. A three-way repeated-measures ANOVA with factors Task (Definition vs. Sound), Condition (Experiment *vs.* Control) and Hemisphere (Left vs. Right ATL) was performed on these values using JASP (JASP team, version .19.3; https://jasp-stats.org/). This analysis highlighted a significant main effect of Task [$F(1,24) = 5.662$,

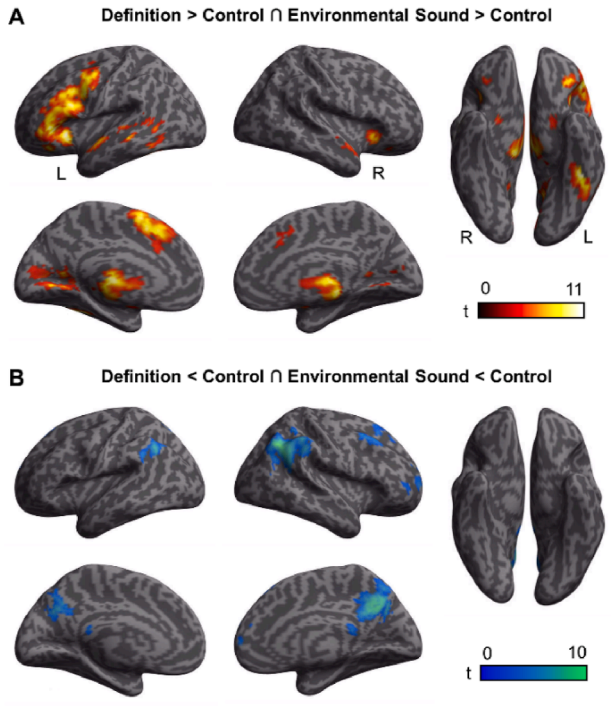

Fig. 4 – A. Conjunction clusters showing significant BOLD signal changes in both Definition > Control and Environmental Sound > Control. B. Conjunction clusters showing significant BOLD signal changes in both Definition < Control and Environmental Sound < Control. A-B results are shown on inflated surface renderings from SPM12. Height thresholded at $p < .001$ with a spatial cluster extent threshold at $p < .05$ (FWE-corrected).

$p = .026$], a significant main effect of Condition ($F = 12.151$, $p = .002$) but no significant main effect of Hemisphere ($F = .019$, $p = .893$). The interaction effect between Task and Condition was also significant ($F = 4.890$, $p = .037$), and the triple interaction effect was trending towards significance ($F = 4.165$, $p = .052$). The other interaction effects were not significant [Task x Hemisphere: $F = .110$, $p = .743$; Condition x Hemisphere: $F = .270$, $p = .608$]. Post-hoc comparisons (Bonferroni-corrected for multiple comparisons) found a significant difference between Experimental and Control stimuli in the naming-to-definitions task in both ATL ROIs [left: $t(24) = 3.947$, $p = .004$; right: $t = 2.899$, $p = .047$]. However, no significant difference between Experimental and Control stimuli was highlighted in the naming-to-environmental-sounds task (left: $t = 1.475$, $p = .919$; right: $t = 2.145$, $p = .254$).

We followed up the null results for the sound contrasts by performing Bayesian paired samples *t*-tests in JASP v0.19.3 (Quintana & Williams, 2018; van Doorn et al., 2021). This analysis demonstrated anecdotal evidence for the null hypothesis (Sound = Control; $BF_{01} = 1.820$ and .684 for the left and right ATL, respectively).

3.2.2. *Conjunction of definitions and environmental sounds*

3.2.2.1. Unrestricted whole-brain analyses. The conjunction analysis investigating common activation in the contrasts

Table 4 – Brain regions showing significant BOLD increases and decreases for the conjunction analysis (null) for contrasts Experimental *vs.* Control.

	Peak MNI (x y z)			Z-score	Cluster size (N voxels)
Definition > Control ∩ Environmental Sounds > Control					
Right ventral caudate nucleus	14	6	0	Inf	11,884
Left lateroventral A37 (FG)	−44	−44	−20	7.22	1356
Left medial A8 (SFG)	−4	8	56	7.13	1608
Left rostral A22 (STG)	−56	−6	−8	6.45	337
Right cerebellar lobule VI	30	−64	−28	6.40	706
Right anterior STS	58	2	−12	4.65	274
Definition < Control ∩ Environmental Sounds < Control					
Right rostroventral A39 (IPL)	48	−60	38	7.52	2868
Right A31 (precuneus)	8	−60	32	6.80	2382
Left caudal A40 (IPL)	−54	−58	36	6.21	579
Left cerebellar lobule Crus I	−46	−74	−38	5.58	233
Right dorsal A9/46 (MFG)	30	24	48	5.17	990
Right A46 (MFG)	26	60	6	4.87	622

Note: Whole-brain analysis, height threshold at $p < .001$ and spatial cluster extent threshold at FWE $p < .05$ (i.e., 274 voxels for Definition > Control ∩ Environmental Sounds > Control; 233 for Definition < Control ∩ Environmental Sounds < Control). Peak locations were determined using the Brainnetome atlas, except for peak clusters in the cerebellum that were located using the CoBrA atlas. Conjunction results are marked by ∩ symbol. FG = fusiform gyrus; IPL = inferior parietal lobule; MFG = middle frontal gyrus; SFG = superior frontal gyrus; STG = superior temporal gyrus; STS = superior temporal sulcus.

Definition > Control ∩ Environmental Sound > Control revealed significant BOLD signal changes in the right ventral caudate nucleus (extending towards the left rostral IFG, A45r), the left lateroventral fusiform gyrus (A37lv), the left medial SFG (A8m), the left rostral STG (A22r), the right anterior STS, and the right cerebellar lobule VI (Fig. 4A and Table 4).

The conjunction analysis with contrasts Definition < Control ∩ Environmental Sound < Control revealed significant BOLD signal changes in the right rostroventral IPL (A39rv, extending towards the right caudal IPL, A40c), the right precuneus (A31), the left caudal IPL (A40c), the right MFG (two clusters: A9/46d and A46), and the left cerebellar lobule Crus I (Fig. 4B and Table 4).

3.2.2.2. SVC analyses. The conjunction analyses with contrasts Definition > Control ∩ Environmental Sound > Control and Definition < Control ∩ Environmental Sound < Control did not reveal significant BOLD signal changes in either vATL ROI using SVC.

3.2.3. *Task-specific activation for naming to definitions*

3.2.3.1. Unrestricted whole-brain analyses. The contrast Definition > Control masked (exclusively) by Environmental Sound > Control revealed task-specific clusters of BOLD signal changes peaking in the left anterior STS, the left dorsomedial POS (in precuneus), the left caudal IPL (A39c), the left rostral IFG (A45r), the left lateral prefrontal thalamus, the left ventrolateral MFG (A8vl), the right rostroventral fusiform gyrus (A20rv), as well as two areas of the cerebellum (right cerebellar lobule Crus I and lobule IX) (Fig. 5A and Table 5).

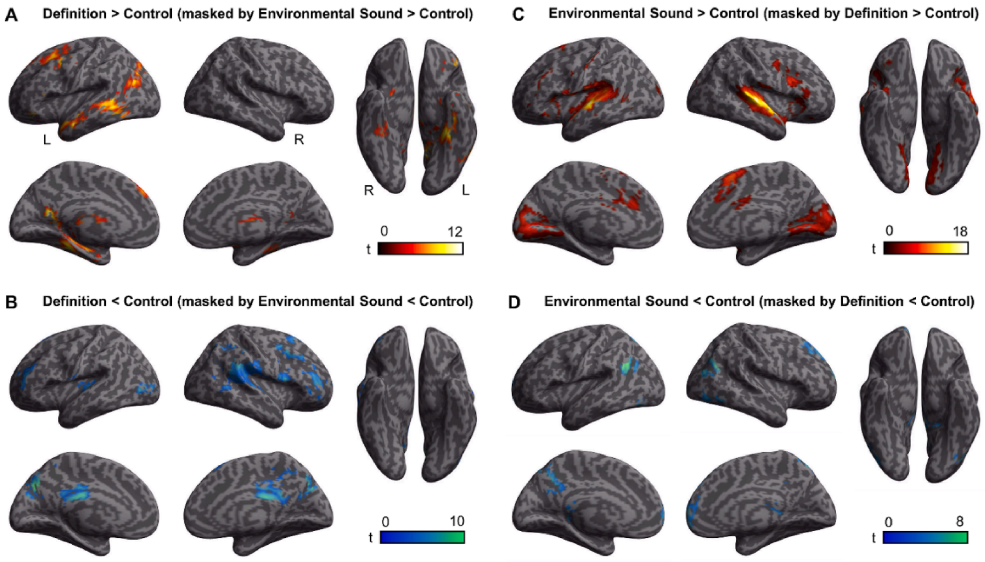

Fig. 5 – A. Significant BOLD signal changes in Definition > Control masked exclusively by Environmental Sound > Control. B. Significant BOLD signal changes in Definition < Control masked exclusively by Environmental Sound < Control. C. Significant BOLD signal changes in Environmental Sound > Control masked exclusively by Definition > Control. D. Significant BOLD signal changes in Environmental Sound < Control masked exclusively by Definition < Control. A-D results are shown on inflated surface renderings from SPM12. Height thresholded at $p < .001$ with a spatial cluster extent threshold at $p < .05$ (FWE-corrected). Mask is set at $p < .001$ (uncorrected, exclusive).

The contrast Definition < Control masked by Environmental Sound < Control revealed task-specific clusters of BOLD signal changes in the left dorsomedial POS (in precuneus), the dorsal cingulate cortex (A23d), the right inferior frontal junction (in MFG), the left V5/MT (in LOC), the left MFG (A46) and the left TE1.0 and TE1.2 (in STG) (Fig. 5B and Table 5).

3.2.3.2. SVC ANALYSES. Significant clusters were detected in the left (peak MNI = −38, −18, −26, cluster size = 57, z-score = 5.58) and right (peak MNI = 34, −10, −34, cluster size = 36, z-score = 4.27) vATL for the contrast Definition > Control exclusively masked by Environmental Sound > Control using SVC. No significant clusters were detected in the vATL ROIs for the reverse contrast Definition < Control masked by Environmental Sound < Control using SVC.

3.2.4. *Task-specific activation for naming to environmental sounds*

3.2.4.1. UNRESTRICTED WHOLE-BRAIN ANALYSES. The contrast Environmental Sound > Control masked (exclusively) by Definition > Control revealed significant task-specific BOLD signal changes peaking in the right caudal STG (A22c, extending anteriorly towards the right TE1.0 and TE1.2), the left TE1.0 and TE1.2 (in STG), the right medial SFG (A8m), and the left rostral cuneus gyrus (Fig. 5C and Table 5).

The contrast Environmental Sound < Control masked by Definition < Control revealed significant task-specific BOLD signal changes peaking in the bilateral rostroventral IPL (A39rv), the right medial orbital gyrus (A14m), the left caudal SPL (A7c), the left caudal hippocampus, and two areas of the cerebellum (bilateral cerebellar lobule Crus I) (Fig. 5D and Table 5).

3.2.4.2. SVC ANALYSES. No significant clusters were observed in the vATL ROIs for either the contrast Environmental Sound > Control masked by Definition > Control or the contrast Environmental Sound < Control masked by Definition < Control using SVC.

4. Discussion

The present study aimed to identify brain regions subtending spoken word production during auditory naming and establish whether vATL regions are commonly activated by linguistic and nonverbal auditory stimuli. Our findings show that both naming tasks activated common regions that are typically reported in the word production literature. These encompassed the left-lateralized inferior frontal and superior temporal areas as well as additional regions that we discuss below. Task-specific activation was also observed, comprising a mostly left-lateralized network for naming to definitions, and a more bilateral network for naming to environmental sounds. While we predicted common activation in the vATL consistent with the proposal of a convergent, crossmodal representation of conceptual knowledge across tasks, we instead observed significant vATL activation only for naming to definitions.

4.1. *Common regions engaged by auditory naming tasks*

The results of the conjunction analysis of Definition > Control ∩ and Environmental Sound > Control revealed common BOLD signal increases in regions typically observed for spoken

Table 5 – Brain regions showing significant task-specific BOLD increases and decreases using exclusive masking for contrasts Experimental *vs.* Control.

	Peak MNI (x y z)			Z-score	Cluster size (N voxels)
Naming to definitions > Control (masked by Naming to environmental sounds > Control)					
Left anterior STS	−54	−6	−18	Inf	903
Left dorsomedial POS (precuneus)	−8	−56	12	7.30	1206
Left caudal A39 (IPL)	−32	−72	30	6.74	2038
Left rostral A45 (IFG)	−56	32	0	5.92	218
Left lateral prefrontal thalamus	−14	−8	12	5.75	973
Left ventrolateral A8 (MFG)	−24	12	52	5.69	1749
Right cerebellar lobule Crus I	24	−78	−26	5.24	367
Right rostroventral A20 (FG)	30	−28	−24	4.99	239
Right cerebellar lobule IX	0	−54	−36	4.94	315
Naming to definitions < Control (masked by Naming to environmental sounds < Control)					
Left dorsomedial POS (precuneus)	−14	−66	34	6.90	4431
Dorsal A23 (cingulate cortex)	0	−22	28	6.31	1099
Right inferior frontal junction (MFG)	44	10	50	5.87	1874
Left V5/MT (LOC)	−48	−80	2	5.33	314
Left A46 (MFG)	−40	56	10	5.33	698
Left TE1.0 and TE1.2 (STG)	−42	−22	8	4.98	490
Naming to environmental sounds > Control (masked by Naming to definitions > Control)					
Right caudal A22 (STG)	64	−20	6	Inf	5769
Left TE1.0 and TE1.2 (STG)	−56	−16	4	Inf	4963
Right medial A8 (SFG)	4	14	56	7.82	1762
Left rostral cuneus gyrus	−2	−86	8	6.59	4597
Naming to environmental sounds < Control (masked by Naming to definitions < Control)					
Left rostroventral A39 (IPL)	−52	−60	30	6.28	442
Left cerebellar lobule Crus I	−20	−88	−30	6.04	368
Right rostroventral A39 (IPL)	56	−64	26	5.82	838
Right medial A14 (orbital gyrus)	2	60	0	5.71	1712
Right cerebellar lobule Crus I	30	−86	−30	5.49	185
Left caudal A7 (SPL)	−4	−74	54	5.35	756
Left caudal hippocampus	−22	−44	10	4.97	516

Note: Whole-brain analysis, height threshold at $p < .001$ and spatial cluster extent threshold at FWE $p < .05$ (i.e., 175 voxels for Definition > Control; 314 for Definition < Control; 897 for Environmental Sounds > Control; 185 for Environmental Sounds < Control). Mask for task-specific activations is set at $p < .001$ (uncorrected mask value, exclusive). Peak locations were determined using the Brainnetome atlas, except for peak clusters in the cerebellum that were located using the CoBrA atlas. FG = fusiform gyrus; IFG = inferior frontal gyrus; IPL = inferior parietal lobule; LOC = lateral occipital cortex; MFG = middle frontal gyrus; POS = parietooccipital sulcus; SFG = superior frontal gyrus; SPL = superior parietal lobule; STG = superior temporal gyrus; STS = superior temporal sulcus.

word production such as the left IFG and STG (e.g., Indefrey, 2011; de Zubicaray, 2023 for reviews supporting their role in phonological and speech monitoring processes). We also observed common activation in regions that are less frequently included in neurobiological models of spoken word production, including the caudate nucleus, the left posterior FG, the left SFG, the right cerebellum and the right anterior STS.

The involvement of the left posterior temporal cortex (FG and/or ITG; also sometimes referred to as the basal temporal language area, see Burnstine et al., 1990; Lüders et al., 1991; Trébuchon-Da Fonseca et al., 2009) has been reported consistently by naming studies across various modalities (e.g., Chouinard & Goodale, 2010; Hamberger et al., 2014; Reilly et al., 2016; Tranel et al., 2003, 2005). Interestingly, a recent fMRI study that used auditory naming to definitions and manipulated the imageability of their definition stimuli and target words (to be named in response to definition stimuli) reported a positive correlation between definition imageability (regardless of target imageability) and activation in the bilateral ventro-temporal cortex, peaking in the left posterior FG (MNI = -42 -52 -16; Garbarini et al., 2020), suggesting that this region is involved in visual imagery evoked by the sentence meaning. A similar interpretation was offered by Tranel et al. (2003) for their posterior ventral temporal activation during naming to environmental sounds. While we lacked imageability ratings for all our definition and sound stimuli, they were matched for concreteness (Brysbaert et al., 2014) which is highly correlated with imageability (e.g., Altarriba et al., 1999) and both stimulus types were rated highly concrete on average (4.76 and 4.75 on a 5-point scale for definitions and sounds, respectively). Hence, a similar visual imagery interpretation seems reasonable for our common posterior fusiform activation. An alternative interpretation for this left FG peak is that it might reflect a top-down activation of the visual orthographic form of the named targets, which would be consistent with coordinates for the so-called Visual Word Form Area (VWFA; e.g., −43, −54, −12 in Talaraich space[7] in Cohen et al., 2000, see also 2002; Ludersdorfer et al., 2016; see also Dębska et al., 2023 for recent review).

In a well-known meta-analysis of 120 functional neuroimaging studies, Binder et al. (2009) reported that the left medial SFG (also called dorsomedial prefrontal cortex, DMPFC) was consistently associated with semantic processing, and others have supported a role for this region in representing conceptual knowledge across modalities (e.g., Fairhall & Caramazza, 2013). More recent studies have refined the role of this region, linking it to controlled, executive aspects of retrieving semantic knowledge (Noonan et al., 2013; Jackson, 2021; Rabini et al., 2023; see also Geranmayeh et al., 2014). Our finding of a common cluster in the left medial SFG provides further evidence supporting this region's involvement in accessing and retrieving semantic knowledge from both verbal and nonverbal auditory inputs.

We also observed significant activation in the anterior STS of both hemispheres, i.e., the dorsal ATL. This finding is consistent with this region receiving unimodal input from auditory association areas and the posterior STG (e.g., Herlin et al., 2021; Jackson et al., 2016; Mesulam, 2023; Pascual et al., 2015). The dorsal ATL is consistently engaged during both passive listening of meaningful auditory stimuli (e.g., Engelien et al., 2006; Humphries et al., 2001; Wilson et al., 2018) and during performance of semantic tasks with auditory stimuli (e.g., Dick et al., 2007; Thierry et al., 2003; Visser and Lambon Ralph 2011). Of note, studies comparing sensory modalities (e.g., picture vs. auditory word or spoken vs.

[7] Conversion to MNI coordinates: −46, −53, −20.

written narratives) indicate that bilateral anterior STS regions may be specifically activated for the auditory rather than other modalities (Visser and Lambon Ralph 2011; Wilson et al., 2018).

Finally, extra-cortical regions such as the cerebellum or basal ganglia (caudate nucleus) have often been reported during speech production (e.g., Abdullaev & Melnichuk, 1997; Ali et al., 2010; Crosson et al., 2003; Riecker et al., 2005; Runnqvist et al., 2016). However, evidence for a production-specific role is equivocal (for a review, see de Zubicaray, 2023). Activity in these regions has instead been attributed to more domain–general processes (e.g., auditory working memory maintenance demands, speech motor planning or language control; Petacchi et al., 2005; Grahn et al., 2008; Ali et al., 2010; Mariën et al., 2014).

The reverse conjunction contrast of Definition < Control ∩ and Environmental Sound < Control revealed significant BOLD signal decreases in bilateral IPL regions, right precuneus, right MFG, and left cerebellum. These clusters are consistent with those reported in an earlier study using covert naming to definitions and a control condition involving passive listening of the same stimuli (Hamberger et al., 2014). Interestingly, these regions overlap with those usually reported as part of the default-mode network in resting state fMRI studies (Buckner, 2012; Buckner & DiNicola, 2019; Raichle, 2015; Raichle et al., 2001). While participants attended and responded to both experimental and control items, the processing demands of the latter were substantially weaker (reversed speech and white noise versus meaningful items) as was their response planning (identical responses to control versus unique responses to meaningful items). This may indicate that the regions implicated in the default mode network respond preferentially to shallow (less meaningful) processing.

4.2. *Task-specific activation*

We also investigated the task-specific correlates of naming to definitions and to environmental sounds via exclusive masking. This revealed multiple and mostly left-lateralized clusters for naming to definitions while naming to environmental sounds was associated with more extensive bilateral activation of the auditory cortices. These differences between input types are likely to reflect the higher demands on linguistic processes for naming to definitions, compared to decoding the idiosyncratic nonverbal, acoustic features of the environmental sounds. In addition, we employed a control condition with reversed speech that comprised identical acoustic information to the sentence definitions, although this was not the case for the white noise control we employed for the environmental sounds. Hence, this difference in acoustic complexity is likely to explain the more extensive bilateral activation we observed in superior temporal regions and auditory cortices for environmental sounds (see Lewis et al., 2004 for discussion on this topic). While white noise is typically used as a control for environmental sounds (e.g., Giraud & Price, 2001; Maeder et al., 2001), future studies might consider opting for a control with better-matched acoustic features, although it should be noted that some reversed sounds can still be recognizable (e.g., Lewis et al., 2004).

4.3. *Task-specific vATL involvement in naming to definitions*

While Definition > Control elicited significant bilateral vATL ROI activation, this was not the case for Environmental Sound > Control. The conjunction analysis that identified common activation across naming to definitions and to environmental sounds likewise did not elicit significant vATL activation, indicating activation in this region was likely specific to the linguistic stimuli. To further explore this relationship, we investigated task-specific activation by exclusively masking each contrast with the other, which again confirmed significant vATL activation only for naming to definitions. We also conducted an independent repeated-measures ANOVA and Bayesian *t*-tests on the beta values extracted from these ATL ROIs which confirmed these results by showing evidence of a significant difference between experimental stimuli and their control in the bilateral ATL for definitions but not for environmental sounds. Overall, these results indicate that the vATL regions were reliably engaged only by the linguistic stimuli in the present study. Note that the lack of significant activation in the vATL ROI for environmental sounds cannot be attributed to signal dropout or distortion as both tasks were acquired with the same multiband sequence in identical participants, and the tSNR in each of the vATL ROIs did not differ significantly between tasks (see Supplementary Material).

As mentioned in the Introduction, several iEEG studies reported gamma modulation to visual naming but not auditory (definition or sound) naming in ventral temporal regions (Ikegaya et al., 2019; Kitazawa et al., 2023; Kojima et al., 2013; Nakai et al., 2019). Yet, using fMRI we found that the vATL was activated during naming to definitions. One reason for the discrepant results might be that the abovementioned iEEG studies did not distinguish between anterior and posterior ventral temporal regions (except for Nakai et al., 2019). Our results align well with another naming study using both intracerebral recordings and fMRI (Forseth et al., 2018) that used more precise anatomical localization of the contacts centered on coordinates (rather than grouping within gyri) and observed common left-lateralized activation for naming to pictures and to auditory definitions in the fusiform gyrus, the y coordinate of which was very similar to the one we used and to those of previous studies investigating semantic processing that labelled this region as anterior temporal (MNI y = −15; Binney et al., 2010). Given the considerable heterogeneity of results over anterior and posterior ventral temporal regions that we noted in our review of the literature, it would be helpful if future studies employed a-priori defined ROIs to distinguish them, especially considering that iEEG methods allow very high spatial and time resolution although with a sparser coverage. In addition, most of the studies mentioned above involved ECoG recordings (except for Ikegaya et al., 2019 who used both ECoG and depth electrodes) which capture electrophysiological activity of neuronal populations in gyri but are much less sensitive to activity in sulci. Further research is thus needed to investigate anterior vs. posterior ventral temporal activity with intracerebral EEG using depth

electrodes that adequately sample electrophysiological activity from both gyri and sulci.

While our vATL ROI was slightly more lateral and superior to the one used by Forseth and colleagues, our results provide further evidence for a role of this ventral region in naming from auditory definitions. However, the fact that we did not observe significant activation in this ROI for naming of environmental sounds is surprising considering the wealth of evidence for a role of the vATL in supporting cross-modal semantic representations of both concrete and abstract concepts (e.g., Binney et al., 2016; Muraki et al., 2025; Rice et al., 2018; Visser and Lambon Ralph 2011). Our findings indicate that environmental sounds do not activate the vATL to the same extent as auditory definitions, consistently with proposals suggesting that they have weaker conceptual-to-lexical connections during naming (e.g., Brownsett et al., 2022; Wöhner et al., 2020).

4.4. *Limitations*

It should be acknowledged that accuracy for naming to environmental sounds was lower (by ~15%) compared to naming to definitions. Despite this imbalance, the number of correct trials per participant for naming environmental sounds generally exceeded 25, which is well within the recommended range of 20–30 trials to observe reliable BOLD signal responses in single-echo fMRI studies (e.g., Huettel & McCarthy, 2001; see also Steele et al., 2016 suggesting that as few as 8 trials could be sufficient depending on the experimental task, sample size, fMRI acquisition parameters, etc.). In addition, we included both error trials and RTs as covariates of no interest in our analyses to control for these effects.

Dick et al. (2016) noted similar brain regions have been observed during spoken language and environmental sound processing across various comprehension tasks. While our study focussed on the role of the vATL in spoken word production, it is worth noting that an fMRI study using a semantic judgment task with pictures, auditory words and environmental sounds as input stimuli reported common activation in the left and right vATL compared to control conditions (Visser and Lambon Ralph 2011). Hence, it is possible the processes required for naming may not activate the vATL to the same extent as a comprehension task (the latter requiring controlled retrieval of specific semantic knowledge), as we noted in the Introduction. In our study, we investigated common as well as task-specific brain activation for two auditory naming tasks using conjunction and exclusive masking approaches (Friston et al., 2005; Nichols et al., 2005) to infer whether the vATL was commonly activated across tasks. However, naming also involves later stages of lexical selection, phonological word form retrieval and articulation (e.g., Indefrey, 2011). Future studies might consider using decoding approaches such as representational similarity analysis (Kriegeskorte et al., 2008) to distinguish the different processing stages underlying naming of auditory input types (e.g., semantic versus phonological). It is also important to mention that, while the control condition used for the naming-to-definitions task was designed to control for low-level auditory and motor aspects of the task, it does not account for the linguistic comprehension processes involved in understanding the definitions. As a result, the observed vATL activation in this contrast may reflect not only conceptual access and production, but also comprehension of the verbal input. In comparison, naming-to-environmental-sounds did not imply understanding lexical stimuli and their relation to each other within a sentence.

Both of our auditory naming tasks comprised target objects from both living (e.g., animals) and non-living categories (e.g., musical instruments), consistent with previous fMRI and TMS studies (e.g., Forseth et al., 2018; Hamberger et al., 2014; Pobric et al., 2010; Reilly et al., 2016). In a PET study, Tranel et al. (2005) reported a cluster in the anterior temporal lobe that responded selectively to animals versus tools during both auditory and visual naming, although they did not observe activation in this region using the same animal sound naming task compared to a low-level baseline auditory task in an earlier PET study (Tranel et al., 2003). Conversely, Pobric et al. (2010) reported that TMS applied to the ATL impaired visual naming of *both* living and non-living items, which they interpreted as evidence for a common semantic hub account. In the present study, 40% (20/50) of the environmental sound stimuli were from the living category, compared to 26% (13/50) of the auditory definitions. However, considering previous evidence of vATL engagement for both living and non-living objects (e.g., Pobric et al., 2010), we consider it unlikely that the differential vATL activity we observed across tasks could be attributed to the different proportions of living versus non-living stimuli. Beyond the living/non-living distinction, we also acknowledge that our stimuli belonged to several semantic categories (similarly to previous studies, e.g., Hamberger et al., 2014; Kitazawa et al., 2023) and that the distribution of target stimuli across these categories differed across tasks (e.g., 11 musical instruments in the environmental sound naming task and none in the definition naming task). Hence, it is possible that the null result we observed for the sound naming task could be due to the different semantic categories involved, although this would imply that the vATL responds in a category-specific manner which is inconsistent with a common semantic hub account (e.g., Lambon Ralph et al., 2017; Pobric et al., 2010). Future studies might consider matching stimuli across tasks according to category and ensure the same target names are elicited.

5. Conclusion

Using auditory naming and multiband fMRI, we investigated brain activity during spoken word production focussing on whether vATL regions are commonly activated by linguistic and nonverbal environmental sound stimuli. Our findings showed that both naming tasks activated common regions that are typically reported in the word production literature. We also observed task-specific activation. While we expected to observe common activation in the vATL consistent with the proposal of a convergent, crossmodal representation of conceptual knowledge across tasks, we observed significant vATL activation only for naming to definitions. These findings suggest a differential pattern of activation in the vATL depending on the type of naming task.

CRediT authorship contribution statement

Angelique Volfart: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katie L. McMahon:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Catherine Liégeois-Chauvel:** Writing – review & editing, Funding acquisition, Conceptualization. **Vitória Piai:** Writing – review & editing, Funding acquisition, Conceptualization. **Greig de Zubicaray:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was supported by the Australian Research Council (grant number DP200100127).

Declaration of competing interest

None.

Acknowledgments

The authors would like to thank all participants who took part in the fMRI experiment and in the online pilot study on Gorilla. This work was supported by the Australian Research Council (grant number DP200100127).

Scientific transparency statement

DATA: Some raw and processed data supporting this research are publicly available, while some are subject to restrictions: https://openneuro.org/datasets/ds005329.
CODE: All analysis code supporting this research is publicly available: https://openneuro.org/datasets/ds005329.
MATERIALS: All study materials supporting this research are publicly available: https://osf.io/fh5xb/
DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.
PRE-REGISTRATION: No part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted. No part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted.
For full details, see the *Scientific Transparency Report* in the supplementary data to the online version of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2025.08.010.

REFERENCES

Abdullaev, Y. G., & Melnichuk, K. V. (1997). Cognitive operations in the human caudate nucleus. *Neuroscience Letters, 234*, 151–155.
Ali, N., Green, D. W., Kherif, F., Devlin, J. T., & Price, C. J. (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience, 22*, 2369–2386.
Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31*, 578–602.
Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage, 13*, 903–919.
Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*, 95–113.
Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26*, 839–851.
Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.
Barsalou, L. W. (2020). *Challenges and opportunities for grounding cognition, 3* (p. 31).
Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*, 2767–2796.
Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon, R. M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex, 20*, 2728–2738.
Binney, R. J., Hoffman, P., & Lambon Ralph, M. A. (2016). Mapping the multiple graded contributions of the anterior temporal lobe representational hub to abstract and social concepts: Evidence from distortion-corrected fMRI. *Cerebral Cortex, 26*, 4227–4241.
Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. In *8th international conference on functional mapping of the human brain, 16*. Sendai, Japan: Available on CD-ROM in NeuroImage. No 2, abstract 497.
Brownsett, S., Mascelloni, M., Gowlett, G., McMahon, K. L., & de Zubicaray, G. (2022). Neighing dogs: Semantic context effects of environmental sounds in spoken word production: A replication and extension. *The Quarterly Journal of Experimental Psychology: QJEP*, Article 17470218221137007.
Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res, 46*, 904–911.
Buckner, R. L. (2012). The serendipitous discovery of the brain's default network. *NeuroImage, 20 YEARS OF fMRI, 62*, 1137–1145.
Buckner, R. L., & DiNicola, L. M. (2019). The brain's default network: Updated anatomy, physiology and evolving insights. *Nature Reviews. Neuroscience, 20*, 593–608.
Burnstine, T. H., Lesser, R. P., Hart, J., Uematsu, S., Zinreich, S. J., Krauss, G. L., Fisher, R. S., Vining, E. P., & Gordon, B. (1990). Characterization of the basal temporal language area in patients with left temporal lobe epilepsy. *Neurology, 40*, 966–970.
Calhoun, V. D., Wager, T. D., Krishnan, A., Rosch, K. S., Seymour, K. E., Nebel, M. B., Mostofsky, S. H., Nyalakanai, P., & Kiehl, K. (2017). The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Human Brain Mapping, 38*, 5331–5342.
Cervenka, M. C., Corines, J., Boatman-Reich, D. F., Eloyan, A., Sheng, X., Franaszczuk, P. J., & Crone, N. E. (2013). Electrocorticographic functional mapping identifies human

cortex critical for auditory and visual naming. *Neuroimage, 69*, 267–276.
Chen, Y.-C., & Spence, C. (2018). Audiovisual semantic interactions between linguistic and nonlinguistic stimuli: The time-courses and categorical specificity. *Journal of Experimental Psychology. Human Perception and Performance, 44*, 1488–1507.
Chen, L., T. Vu, A., Xu, J., Moeller, S., Ugurbil, K., Yacoub, E., & Feinberg, D. A. (2015). Evaluation of highly accelerated simultaneous multi-slice EPI for fMRI. *Neuroimage, 104*, 452–459.
Chouinard, P. A., & Goodale, M. A. (2010). Category-specific neural processing for naming pictures of animals and naming pictures of tools: An ALE meta-analysis. *Neuropsychologia, 48*, 409–418.
Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain: a Journal of Neurology, 123*, 291–307.
Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain: a Journal of Neurology, 125*, 1054–1069.
Crosson, B., Benefield, H., Cato, M. A., Sadek, J. R., Moore, A. B., Wierenga, C. E., Gopinath, K., Soltysik, D., Bauer, R. M., Auerbach, E. J., Gökçay, D., Leonard, C. M., & Briggs, R. W. (2003). Left and right basal ganglia and frontal activity during language generation: Contributions to lexical, semantic, and phonological processes. *Journal of the International Neuropsychological Society: JINS, 9*, 1061–1077.
Dębska, A., Wójcik, M., Chyl, K., Dzięgiel-Fivet, G., & Jednoróg, K. (2023). Beyond the Visual Word Form Area – a cognitive characterization of the left ventral occipitotemporal cortex. *Front Hum Neurosci, 17*.
de Zubicaray, G. I. (2023). The neural organisation of language production: Evidence from neuroimaging and neuromodulation. In R. J. Hartsuiker, & K. Strijkers (Eds.), *Language production. Current issues in the psychology of language*. Routledge.
Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.
Demetriou, L., Kowalczyk, O. S., Tyson, G., Bello, T., Newbould, R. D., & Wall, M. B. (2018). A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *Neuroimage, 176*, 404–416.
Dick, F., Krishnan, S., Leech, R., & Saygin, A. P. (2016). Chapter 89 - environmental sounds. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 1121–1138). San Diego: Academic Press.
Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., Wilson, S., Bates, E., & Pizzamiglio, L. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: Evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience, 19*, 799–816.
Ding, J., Chen, K., Chen, Y., Fang, Y., Yang, Q., Lv, Y., Lin, N., Bi, Y., Guo, Q., & Han, Z. (2016). The left fusiform gyrus is a critical region contributing to the core behavioral profile of semantic dementia. *Front Hum Neurosci, 10*.
Embleton, K. V., Haroon, H. A., Morris, D. M., Ralph, M. A. L., & Parker, G. J. M. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping, 31*, 1570–1587.
Engelien, A., Tüscher, O., Hermans, W., Isenberg, N., Eidelberg, D., Frith, C., Stern, E., & Silbersweig, D. (2006). Functional neuroanatomy of non-verbal semantic sound processing in humans. *Journal of Neural Transmission, 113*, 599–608.
Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 33*, 10552–10558.
Fairs, A., Fargier, R., & Strijkers, K. (2022). *Shared or different: How linked are word production and comprehension? TIPA Travaux interdisciplinaires sur la parole et le langage*.
Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex, 26*, 3508–3526.
Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *Plos One, 5*, Article e15710.
Forseth, K. J., Kadipasaoglu, C. M., Conner, C. R., Hickok, G., Knight, R. T., & Tandon, N. (2018). A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain: a Journal of Neurology, 141*, 2112–2126.
Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *Neuroimage, 25*, 661–667.
Garbarini, F., Calzavarini, F., Diano, M., Biggio, M., Barbero, C., Radicioni, D. P., Geminiani, G., Sacco, K., & Marconi, D. (2020). Imageability effect on the functional brain activity during a naming to definition task. *Neuropsychologia, 137*, Article 107275.
Geranmayeh, F., Wise, R. J. S., Mehta, A., & Leech, R. (2014). Overlapping networks engaged during spoken language production and its cognitive control. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 34*, 8728–8740.
Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience, 13*, 754–765.
Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology. General, 118*, 13–42.
Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology, 86*, 141–155.
Hamberger, M. J., Goodman, R. R., Perrine, K., & Tamny, T. (2001). Anatomic dissociation of auditory and visual naming in the lateral temporal cortex. *Neurology, 56*, 56–61.
Hamberger, M. J., Habeck, C. G., Pantazatos, S. P., Williams, A. C., & Hirsch, J. (2014). Shared space, separate processes: Neural activation patterns for auditory description and visual object naming in healthy adults. *Hum Brain Mapp, 35*, 2507–2520.
Hamberger, M. J., Heydari, N., Caccappolo, E., & Seidel, W. T. (2021). Naming in older adults: Complementary auditory and visual assessment. *Journal of the International Neuropsychological Society*, 1–14.
Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society, 9*, 479–489.
Herlin, B., Navarro, V., & Dupont, S. (2021). The temporal pole: From anatomy to function—a literature appraisal. *Journal of Chemical Neuroanatomy, 113*, Article 101925.
Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for environmental sound stimuli. *Plos One, 8*, Article e73382.
Hoffman, P., & Lambon, R. M. A. (2018). From percept to concept in the ventral temporal lobes: Graded hemispheric specialisation based on stimulus and task. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 101*, 107–118.

Huettel, S. A., & McCarthy, G. (2001). The effects of single-trial averaging upon the spatial extent of fMRI activation. *Neuroreport, 12*, 2411.
Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *Neuroreport, 12*, 1749–1752.
Ikegaya, N., Motoi, H., Iijima, K., Takayama, Y., Kambara, T., Sugiura, A., Silverstein, B. H., Iwasaki, M., & Asano, E. (2019). Spatiotemporal dynamics of auditory and picture naming-related high-gamma modulations: A study of Japanese-speaking patients. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 130*, 1446–1454.
Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Front Psychol, 2*, 255.
Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.
Jackson, R. L. (2021). The neural correlates of semantic control revisited. *Neuroimage, 224*, Article 117444.
Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 36*, 1490–1501.
Kemmerer, D. (2019). From blueprints to brain maps: The status of the lemma model in cognitive neuroscience. *Language, Cognition and Neuroscience, 34*, 1085–1116.
Kitazawa, Y., Sonoda, M., Sakakura, K., Mitsuhashi, T., Firestone, E., Ueda, R., Kambara, T., Iwaki, H., Luat, A. F., Marupudi, N. I., Sood, S., & Asano, E. (2023). Intra– and inter–hemispheric network dynamics supporting object recognition and speech production. *Neuroimage, 270*, Article 119954.
Kojima, K., Brown, E. C., Matsuzaki, N., Rothermel, R., Fuerst, D., Shah, A., Mittal, S., Sood, S., & Asano, E. (2013). Gamma activity modulated by picture and auditory naming tasks: Intracranial recording in patients with focal epilepsy. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 124*, 1737–1744.
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci, 2*.
Lüders, H., Lesser, R. P., Hahn, J. F., Dinner, D. S., Morris, H. H., Wyllie, E., & Godoy, J. (1991). Basal temporal language area. *Brain: a Journal of Neurology, 114*, 743–754.
Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews. Neuroscience, 18*, 42–55.
Levelt, W. J. M. (1999). Models of word production. *Trends Cogn Sci., 3*, 223–232.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.
Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14*, 1008–1021.
Ludersdorfer, P., Wimmer, H., Richlan, F., Schurz, M., Hutzler, F., & Kronbichler, M. (2016). Left ventral occipitotemporal activation during orthographic and semantic processing of auditory words. *Neuroimage, 124*, 834–842.
Macey, P. M., Macey, K. E., Kumar, R., & Harper, R. M. (2004). A method for removal of global effects from fMRI time series. *Neuroimage, 22*, 360–366.
Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.-P., Pittet, A., & Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage, 14*, 802–816.
Malow, B. A., Blaxton, T. A., Sato, S., Bookheimer, S. Y., Kufta, C. V., Figlozzi, C. M., & Theodore, W. H. (1996). Cortical stimulation elicits regional distinctions in auditory and visual naming. *Epilepsia, 37*, 245–252.
Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology, 22*, 830–864.
Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C. H. S., Beaton, A., Desmond, J., De Witte, E., Fawcett, A. J., Hertrich, I., Küper, M., Leggio, M., Marvel, C., Molinari, M., Murdoch, B. E., Nicolson, R. I., Schmahmann, J. D., Stoodley, C. J., Thürling, M., Timmann, D., Wouters, E., & Ziegler, W. (2014). Consensus paper: Language and the cerebellum: An ongoing enigma. *Cerebellum, 13*, 386–410.
Mehta, S., Grabowski, T. J., Razavi, M., Eaton, B., & Bolinger, L. (2006). Analysis of speech-related variance in rapid event-related fMRI using a time-aware acquisition system. *Neuroimage, 29*, 1278–1293.
Mesulam, M.-M. (2023). Temporopolar regions of the human brain. *Brain, 146*(1), 20–41.
Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., Fryer, T. D., Williams, G. B., Hodges, J. R., & Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain: a Journal of Neurology, 133*, 3256–3268.
Muraki, E. J., Pexman, P. M., & Binney, R. J. (2025). Mapping contributions of the anterior temporal semantic hub to the processing of abstract and concrete verbs. *Human Brain Mapping, 46*, Article e70210.
Nakai, Y., Sugiura, A., Brown, E. C., Sonoda, M., Jeong, J.-W., Rothermel, R., Luat, A. F., Sood, S., & Asano, E. (2019). Four-dimensional functional cortical maps of visual and auditory language: Intracranial recording. *Epilepsia, 60*, 255–267.
Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25*, 653–660.
Nieto-Castañón, A. (2022). *Preparing fMRI data for statistical analysis*.
Nieto-Castañón, A., & Whitfield-Gabrieli, S. (2022). *CONN functional connectivity toolbox: Rrid SCR_009550*.
Noonan, K. A., Jefferies, E., Visser, M., & Lambon, R. M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: Evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J Cogn Neurosci, 25*, 1824–1850.
Nozari, N. (2022). The neural basis of word production. In A. Papafragou, J. C. Trueswell, & L. R. Gleitman (Eds.), *The oxford handbook of the mental lexicon* (pp. 536–558). New York: Oxford University Press.
Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., & Conturo, T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage, 6*, 156–167.
Papademetris, X., Jackowski, M. P., Rajeevan, N., DiStasio, M., Okuda, H., Constable, R. T., & Staib, L. H. (2006). BioImage Suite: An integrated medical image analysis suite: An update. *Insight J, 2006*, 209.
Park, M. T. M., Pipitone, J., Baer, L. H., Winterburn, J. L., Shah, Y., Chavez, S., Schira, M. M., Lobaugh, N. J., Lerch, J. P., Voineskos, A. N., & Chakravarty, M. M. (2014). Derivation of high-resolution MRI atlases of the human cerebellum at 3 T and segmentation using multiple automatically generated templates. *Neuroimage, 95*, 217–231.

Pascual, B., Masdeu, J. C., Hollenbeck, M., Makris, N., Insausti, R., Ding, S.-L., & Dickerson, B. C. (2015). Large-scale brain networks of the human left temporal pole: A functional connectivity MRI study. *Cerebral Cortex, 25*, 680–702.

Patel, T., Morales, M., Pickering, M. J., & Hoffman, P. (2023). A common neural code for meaning in discourse production and comprehension. *Neuroimage, 279*, Article 120295.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience, 8*, 976–987.

Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25*, 118–128.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology: CB, 20*, 964–968.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage, 84*, 320–341.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage, 20 YEARS OF fMRI, 62*, 816–847.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 18*, 178.

Rabini, G., Ubaldi, S., & Fairhall, S. L. (2023). Task-based activation and resting-state connectivity predict individual differences in semantic capacity for complex semantic knowledge. *Commun Biol, 6*, 1–13.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). *A default mode of brain function, 98* (pp. 676–682). Proceedings of the National Academy of Sciences.

Reilly, J., Garcia, A., & Binney, R. J. (2016). Does the sound of a barking dog activate its corresponding visual form? An fMRI investigation of modality-specific semantic access. *Brain and Language, 159*, 45–59.

Rice, G. E., Hoffman, P., Binney, R. J., & Lambon Ralph, M. A. (2018). Concrete versus abstract forms of social concept: An fMRI comparison of knowledge about people versus social terms. *Philos Trans R Soc Lond B Biol Sci., 373*, Article 20170136.

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology, 64*, 700–706.

Roos, N. M., Takashima, A., & Piai, V. (2023). Functional neuroanatomy of lexical access in contextually and visually guided spoken word production. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 159*, 254–267.

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*, 1864–1881.

Runnqvist, E., Bonnard, M., Gauvin, H. S., Attarian, S., Trébuchon, A., Hartsuiker, R. J., & Alario, F.-X. (2016). Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 81*, 203–214.

Steele, V. R., Anderson, N. E., Claus, E. D., Bernat, E. M., Rao, V., Assaf, M., Pearlson, G. D., Calhoun, V. D., & Kiehl, K. A. (2016). Neuroimaging measures of error-processing: Extracting reliable signals from event-related potentials and functional magnetic resonance imaging. *Neuroimage, 132*, 247–260.

Thierry, G., Giraud, A. L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron, 38*, 499–506.

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S. A., & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. *Neurobiology of Language*, 1–21.

Tomasino, B., Canderan, C., Marin, D., Maieron, M., Gremese, M., D'Agostini, S., Fabbro, F., & Skrap, M. (2015). Identifying environmental sounds: A multimodal mapping study. *Frontiers in Human Neuroscience., 9*.

Trébuchon-Da Fonseca, A., Guedj, E., Alario, F.-X., Laguitton, V., Mundler, O., Chauvel, P., & Liegeois-Chauvel, C. (2009). Brain regions underlying word finding difficulties in temporal lobe epilepsy. *Brain: a Journal of Neurology, 132*, 2772–2784.

Tranel, D., Damasio, H., Eichhorn, G. R., Grabowski, T., Ponto, L. L. B., & Hichwa, R. D. (2003). Neural correlates of naming animals from their characteristic sounds. *Neuropsychologia, 41*, 847–854.

Tranel, D., Grabowski, T. J., Lyon, J., & Damasio, H. (2005). Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *J Cogn Neurosci, 17*, 1293–1305.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*, 813–826.

van Scherpenberg, C., Just, A. C., & Hauber, R. C. (2020). *Check voice onset times from Chronset with Praat script* [WWW Document]. URL https://osf.io/fmwqb/.

Visser, M., Jefferies, E., Embleton, K. V., & Lambon, R. M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J Cogn Neurosci, 24*, 1766–1778.

Visser, M., & Lambon, R. M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J Cogn Neurosci, 23*, 3121–3131.

Volfart, A., McMahon, K. L., & de Zubicaray, G. I. (2024). A comparison of denoising approaches for spoken word production related artefacts in continuous multiband fMRI data. *Neurobiology of Language*, 1–45.

Wöhner, S., Jescheniak, J. D., & Mädebach, A. (2020). Semantic interference is not modality specific: Evidence from sound naming with distractor pictures. *The Quarterly Journal of Experimental Psychology: QJEP, 73*, 2290–2308.

Walenski, M., Europa, E., Caplan, D., & Thompson, C. K. (2019). Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. *Human Brain Mapping, 40*, 2275–2304.

Wall, M. B. (2023). Multiband acquisition sequences for fMRI: Proceed with caution. *Aperture Neuro, 3*.

Whitfield-Gabrieli, S., Nieto-Castañón, A., & Ghosh, S. (2011). *Artifact detection tools (ART)*.

Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage, 171*, 62–74.
Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping, 4*, 58–73.
Yochim, B. P., Beaudreau, S. A., Fairchild, J. K., Yutsis, M. V., Raymond, N., Friedman, L., & Yesavage, J. (2015). Verbal naming test for use with older adults: Development and initial validation. *Journal of the International Neuropsychological Society, 21*, 239–248.