NeuroImage 248 (2022) 118811

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Speech naturalness detection and language representation in the dog brain

Laura V. Cuaya[a,b,*], Raúl Hernández-Pérez[a,b], Marianna Boros[a,b], Andrea Deme[c,d], Attila Andics[a,b,*]

[a] Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary
[b] MTA-ELTE 'Lendület' Neuroethology of Communication Research Group, Hungarian Academy of Sciences – Eötvös Loránd University, Budapest, Hungary
[c] Department of Applied Linguistics and Phonetics, Faculty of Humanities, Eötvös Loránd University, Budapest, Hungary
[d] MTA-ELTE 'Lendület' Lingual Articulation Research Group, Budapest, Hungary

ARTICLE INFO

Keywords:
Speech detection
Language representation
Dogs
Functional magnetic resonance imaging (fMRI)

ABSTRACT

Family dogs are exposed to a continuous flow of human speech throughout their lives. However, the extent of their abilities in speech perception is unknown. Here, we used functional magnetic resonance imaging (fMRI) to test speech detection and language representation in the dog brain. Dogs (n = 18) listened to natural speech and scrambled speech in a familiar and an unfamiliar language. Speech scrambling distorts auditory regularities specific to speech and to a given language, but keeps spectral voice cues intact. We hypothesized that if dogs can extract auditory regularities of speech, and of a familiar language, then there will be distinct patterns of brain activity for natural speech vs. scrambled speech, and also for familiar vs. unfamiliar language. Using multivoxel pattern analysis (MVPA) we found that bilateral auditory cortical regions represented natural speech and scrambled speech differently; with a better classifier performance in longer-headed dogs in a right auditory region. This neural capacity for speech detection was not based on preferential processing for speech but rather on sensitivity to sound naturalness. Furthermore, in case of natural speech, distinct activity patterns were found for the two languages in the secondary auditory cortex and in the precruciate gyrus; with a greater difference in responses to the familiar and unfamiliar languages in older dogs, indicating a role for the amount of language exposure. No regions represented differently the scrambled versions of the two languages, suggesting that the activity difference between languages in natural speech reflected sensitivity to language-specific regularities rather than to spectral voice cues. These findings suggest that separate cortical regions support speech naturalness detection and language representation in the dog brain.

1. Introduction

Every language is characterized by acoustic regularities, such as prosodic features or the distribution of speech sounds, that humans learn about, well before the semantically or syntactically informed phases of language acquisition. At birth, humans are already capable of discriminating speech from similarly complex non-speech stimuli (Dehaene-Lambertz et al., 2002; Ramus et al., 2000; Vouloumanos and Werker, 2007; Vouloumanos et al., 2004). Infants also discriminate familiar from unfamiliar languages belonging to different rhythm classes, while discrimination between two languages of the same rhythm class appears to require previous familiarization with one of the two (Bosch and Sebastián-Gallés, 1997; Nazzi et al., 1998; Nazzi et al., 2000). These capacities of preverbal infants suggest that the processes underlying speech detection and language discrimination do not require higher level linguistic competence (Kuhl, 1994; Rosen, 1992; Vouloumanos et al., 2010), but may entail computations on and learning about low-level features that may also be present in other species.

Indeed, discrimination between speech and non-speech stimuli, as well as between languages has also been demonstrated in non-human species. Using functional magnetic resonance imaging (fMRI) Joly and colleagues (Joly et al., 2012) found that in macaques speech elicits stronger activity than scrambled speech in the lateral belt and parabelt regions. In behavioural studies, cotton-top tamarin monkeys showed an ability to discriminate languages without previous training (Ramus et al., 2000). Moreover, rats and even Java sparrows have been shown to discriminate between languages after training, and they were even able to generalize this ability to new utterances from the same languages (Toro and Trobalon, 2003; Toro and Trobalon, 2005; Watanabe et al., 2006).

Similarly to infants, neither monkeys, nor rats or birds were able to discriminate between speech stimuli from different languages played

* Corresponding authors at: Department of Ethology, Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/C, Hungary.
E-mail addresses: lauravcuaya@gmail.com (L.V. Cuaya), attila.andics@gmail.com (A. Andics).

https://doi.org/10.1016/j.neuroimage.2021.118811.
Received 22 October 2021; Received in revised form 8 December 2021; Accepted 10 December 2021
Available online 12 December 2021.

backwards, which suggests that language discrimination in all these species is based on properties unique to speech, and is not merely an ability to discriminate between complex auditory stimuli in general (Ramus et al., 2000; Toro and Trobalon, 2003; JM Toro and Trobalon, 2005; Watanabe et al., 2006).

Functionally, speech detection is dependent on the processing of precise temporal arrangement of spectro-temporal features (Price et al., 2005), while language discrimination requires learning about auditory regularities (e.g. speech sound inventories, syllable structure, stress pattern, pitch-related characteristics) that characterize a given language. Neuroimaging studies revealed a central role of the superior temporal cortex for both of these processes in human adults (Joly et al., 2012; Belin et al., 2002; Belin et al., 2000; Hickok and Poeppel, 2000; Overath et al., 2015; Okada et al., 2010; Price, 2012; Zhao et al., 2008).

Dogs kept as companion animals share with humans an intense exposure to speech in their natural environment (Pongrácz et al., 2001). Human voices in general, and speech in particular, are not only familiar, but also highly relevant to dogs. This makes dogs a useful comparison species for exploring the evolutionary bases of human voice and speech perception (Andics and Miklósi, 2018; Andics and Faragó, 2018). Behavioural evidence suggests that dogs are sensitive to both segmental and suprasegmental cues in speech (Ratcliffe and Reby, 2014; Root-gutteridge et al., 2019). Recent neuroimaging findings show that dogs can make use of these cues to process speaker identity (Boros et al., 2020), emotional prosody and word familiarity (Andics et al., 2016) or even word meaningfulness (Prichard et al., 2018). Notably, as evidenced by behavioural studies, there are at least a few exceptional dogs that have a large vocabulary and learn new words referencing objects rapidly (Pilley and Reid, 2011; Pilley, 2013; Kaminski et al., 2004).

Whereas fMRI studies indicate the involvement of the dog temporal cortex in the processing of human speech, there has been no evidence to date that dog brains can discriminate speech from non-speech stimuli. Additionally, while dogs, like humans, are typically over-exposed to a specific language, it has remained unexplored, both behaviourally and neurally, if they are able to extract language-specific auditory regularities from speech and distinguish a familiar language from an unfamiliar one.

To test dog brains' capacity for speech detection and language representation, and to reveal the neural processes involved, here we presented awake dogs with natural and scrambled speech (using a quilting algorithm, see (Overath et al., 2015)) of a familiar and an unfamiliar language (Hungarian and Spanish) during an fMRI test. We used multivariate pattern analysis (MVPA), hypothesizing that speech detection (i.e. a prelinguistic, acoustic analysis of speech) will be reflected by differential activity patterns for speech vs. scrambled speech, and language representation (i.e. sensitivity to language-specific regularities in the speech signal) will be reflected in differential activity patterns to speech in a familiar vs. an unfamiliar language.

2. Methods

2.1. Participants

Eighteen adult family dogs (nine females; aged between 3 and 11 years old; mean age = 6.6, SD = 2.7; five golden retrievers, six border collies, two Australian shepherds, one labradoodle, one cocker spaniel and three mixed breeds) participated in the study. All participants were trained previously to remain still inside an MRI scanner. All experimental procedures were approved by the National Animal Experimentation Ethics Committee (number of ethical permission: PEI/001/1490–4/2015). Owners volunteered for the study and didn't receive monetary compensation. Participants could leave the sessions at any time.

Additionally, sixteen adult humans participated in an online survey to rate the naturalness of the stimuli (mean age = 30.9 years, range 25 – 38 years, seven males) without knowledge of either Hungarian or Spanish (native languages: four French, three English, two Hebrew, two Italian, two Polish, one German, one Portuguese, and one Swedish speaker). All participants provided informed consent to participate in the survey, which was carried out in accordance with the relevant guidelines and with the approval of the Institutional Review Board of the Institute of Biology, Eötvös Loránd University, Budapest, Hungary (reference number of ethical permission: 2019/49).

2.2. Stimuli

We used speech samples from Hungarian and Spanish. Hungarian was the language spoken in the environment of 16 dogs, Spanish of 2 dogs (familiar language); the other language was unfamiliar to all dogs (unfamiliar language). Although Hungarian and Spanish differ both segmentally (e.g., speech sound inventories overlap only partially, (International Phonetic Association IPAS 1999; Martínez, 2011)) and suprasegmentally (e.g., dominant stress patterns are different, (Lleó, 2003; Siptár and Törkenczy, 2000)), both languages belong to the same rhythm class (Siptár and Törkenczy, 2000; Kohári, 2018; Nespor et al., 2011), based on objective rhythm measures (i.e., proportion of vocalic intervals and variability of consonant intervals).

Our linguistic material consisted of a recording of the XXI chapter of *The Little Prince* written by Antoine de Saint-Exupéry read by two different native, female speakers, with similar timbre, and vocal characteristics (see below), one in each language. Our choice for one speaker per language was motivated by previous infant studies which showed that increasing speaker variability impairs language discrimination performance (Ramus et al., 2000; Jusczyk et al., 1992). The text, as well as the speakers were unknown to all dogs and the text was recorded with a lively, engaging intonation. Since this is the first study that explored dogs' speech detection abilities in general, we decided to use natural stimuli here, similarly to neuroimaging studies with infants (Dehaene-Lambertz et al., 2002; Homae et al., 2006), even if this limited the possibility to systematically test for specific auditory cues which might be detected by dogs.

Stimuli were created by extracting full sentences from each recording (all started and ended with 0.3 s of silence), which were concatenated into 8.3-second-long fragments. Thus, each trial contained several sentences, and always began and ended at a sentence boundary. With this method, we obtained 24 unique speech fragments of each language. There were no significant differences in the number of syllables per fragment across languages (Hungarian: $M = 41$, SD = 3.3, range 36 to 49; Spanish: $M = 40$, SD = 3.2, range 32 to 46; $t(46) = -1.52$, $p = 0.14$). We also quantified and compared some speaker-related differences in the samples. First, we calculated the mean fundamental frequency (f0) as measured and averaged throughout the total duration of each speech fragment. Second, in the same time window, we measured the first five formants as a mean of the long-term average spectra, and calculated the average spacing between them, i.e. formant dispersion (FD) (Riede and Fitch, 1999). These parameters did not differ across speakers (f0: Hungarian speaker: $M = 197$ Hz, SD = 8.5 Hz, range 179 to 213 Hz; Spanish speaker: $M = 202$ Hz, SD = 13.2 Hz, range 176 to 216 Hz; $t(46) = -1.605$, $p = 0.12$. FD: Hungarian speaker: $M = 1056$ Hz, SD = 12 Hz, range 1029 to 1079; Spanish speaker: $M = 1059$ Hz, SD = 12 Hz, range 1037 to 1092; $t(46) = -0.8$, $p = 0.21$).

As a control condition for speech, we created scrambled fragments using the quilting algorithm (Overath et al., 2015), which extracted 30 ms slices of the speech fragments mentioned above, and then concatenated these slices in a random order while minimizing the acoustic artifacts resulting from concatenation. In quilted speech stimuli, low-level acoustic properties (e.g. mean fundamental frequency, total signal duration) are maintained, but higher-level information (e.g. most of the segmental and morphological content, and the prosodic features, i.e. intonation, stress, accent, rhythm) is disrupted. Behavioural and neuroimaging studies show that disrupting the temporal organization of speech on different time scales can seriously impact speech detection, and thus quilt manipulation provides a suitable control condition in the

examination of speech detection ability (Overath et al., 2015; Norman-Haignere et al., 2015; Overath and Lee, 2018; Saberi and Perrott, 1999). With this method, we obtained 24 unique scrambled speech fragments of each language. Stimuli were equalized at 68 dB and digitized at 16 bit/32 kHz. Similarly to (Overath et al., 2015) we acquired naturalness ratings for each stimulus from the human participants on a 7-point Likert scale using an online survey.

2.3. Design

We used a block design, each condition block lasted 20 s and consisted of two consecutive fragments (8.3 s each) with a no-stimulus gap of 1.7 s for data acquisition. Designs with similar stimulus durations have been used successfully with infants (Dehaene-Lambertz et al., 2002; Nazzi et al., 2000). Four condition blocks were created: Natural speech in a Familiar language (NF), Natural speech in an Unfamiliar language (NU), Scrambled speech in a Familiar language (SF), and Scrambled speech in an Unfamiliar language (SU). Each condition block was presented three times in a pseudo-random order (two blocks of the same condition were never presented consecutively), and three additional silence blocks were interspersed, amounting to 15 blocks per run in total. Four different runs were created, each with different fragments. The order of runs was counterbalanced between participants.

2.4. Data acquisition

We used a sparse sampling acquisition in a 3 T Philips Ingenia scanner with an eight-channel dStream Pediatric Torso coil. Dogs were fitted with noise-protecting earmuffs which were used to present the stimuli during the silent gaps of the sparse sampling protocol. Blood-oxygen-level dependent (BOLD) images of the whole-brain were acquired with a gradient-echo-echo-planar imaging (EPI) sequence (40 transverse slices, 2 mm thickness, 0.5 mm gap; TR = 10 s (1.680 s for acquisition); TE = 12 ms; flip angle = 90°; acquisition matrix 80 × 58; spatial resolution 2.5 mm × 2.5 mm; 32 vol and 1 dummy scan). Sparse sampling acquisition with similar parameters has been well established by previous human (Belin et al., 2000; Perrachione and Satrajit, 2013) and dog (Boros et al., 2020; Andics et al., 2014) studies. During the acquisition, the trainer and the owner remained inside the scanner room. The acquisition took place in different sessions until each participant reached four runs in total with a maximum movement of 3 mm in any direction and less than 1º rotation in any direction (mean framewise displacement (Power et al., 2012) across all dogs and runs = 0.38 mm, the proportion of FD > 1 mm volumes was 6.6%). One dog only completed three runs.

2.5. Data analysis

Raw functional images were pre-processed using FSL 5.0.11. Because automatic tools commonly used to process human images are not optimal to handle dog images during pre-processing, some of the steps we followed were manual. We first reoriented all the functional images to match the dog template (Czeibert et al., 2019) using FSLUTILS. We then calculated a mean functional image by aligning each run to the first volume of the run using FSL's FLIRT (Jenkinson et al., 2012) and then averaged all volumes for each dog. Each run was then aligned to the image using FLIRT. We then skull-stripped all runs using a binary mask manually drawn over the mean functional image of each run. The mean functional image was manually coregistered to the dog's own structural image using 3D Slicer. The resulting image was then manually transformed to the dog template (Czeibert et al., 2019) using Amira 3D, thus creating a mean normalized image (in the dog template space) for each dog. We calculated a transformation matrix from the mean image to the mean normalized image using FLIRT and applied it to each previously aligned run. The images were then smoothed using a Gaussian kernel (5 mm FWHM).

We ran a whole-brain GLM analysis using SPM12. Statistical parametric maps were generated using the linear combination of functions derived by convolving the standard SPM hemodynamic response function (HRF) with the time series of stimulus categories (NF, NU, SF, SU, silence). We decided not to censor out high FD volumes (> 1 mm) but accounted for all head motion by adding a separate regressor of no interest for each of the six motion parameters. Individual contrast images were computed for all sounds versus silence, and for each acoustic condition versus silence. All results are reported at an uncorrected voxel threshold of $p < 0.001$, and FWE-corrected at cluster level ($p < 0.05$).

We used MVPA to reveal whether there are distinct cerebral patterns for natural vs. scrambled speech and familiar vs. unfamiliar language in the dog brain. Using FSL 5.0.11, for each dog, runs were motion corrected and spatially aligned to the first volume of the first run. Each run was filtered using high-pass filter to remove low-frequency signals. We carried out MVPA using the PyMVPA software package (Hanke et al., 2009) and the LibSVM's implementation of the linear support vector machine classifier (LSVM www.csie.ntu.edu.tw/~cjlin/libsvm). The time series were labelled according to the onsets of the events during the run. The events were modelled to account for the peak of the HRF. Each acquisition was linearly detrended and z-scored.

To assess how certain subregions of the temporal cortex contributed to encoding information about a particular stimulus, for each participant we created a classification map using MVPA within the brain using a searchlight approach (Kriegeskorte et al., 2006). That is, we took a voxel within the region and created a sphere (radius = 3 voxels) around it. All the voxels within the sphere were considered in the analysis. All but one run was used to train a two-way LSVM classifier, the remaining run was used to test the classifier. We repeated this process following a cross validation scheme in which each run was used once as test. We then calculated the mean classifier accuracy and projected it back to the center of the sphere; this process was repeated for every voxel in the brain, thus creating a classification map for each participant.

To assess at the group level whether a sphere of voxels around a particular voxel consistently encoded information about a stimulus, we performed a one-sample *t*-test using the performance of all participants for the voxels within the sphere surrounding the voxel. The t-value was projected back to the voxel, thus creating a group result map. We tested for speech detection, i.e. Natural speech vs. Scrambled speech (NF+NU vs. SF+SU), irrespective of language familiarity; and for language representation, overall, i.e. Familiar language vs. Unfamiliar language (NF+SF vs. NU+SU), specifically for Natural speech (NF vs. NU), and finally, as control, for Scrambled speech (SF vs. SU). To quantify effect size, we calculated Cohen's *d* values for the classifier performances measured (Cohen, 1994; Wilson et al., 2020).

To estimate the cluster size expected by chance, we used a procedure of cluster size control (Stelzer et al., 2013) by calculating an accuracy map for each participant under a no-signal condition, this is, we randomly shuffled all stimulus labels for each run. We then created an accuracy map by creating a sphere around a voxel and using all the voxels within the sphere to train and test a LSVM classifier in a leave-one-out cross-validation scheme. We repeated this procedure for all the voxels and created a no-signal accuracy map. We followed the same procedure for all participants and calculated a one-sample *t*-test (expected mean of 0.5) on each voxel using as samples the corresponding accuracies for all participants and assigned the result of the *t*-test to the voxel, thus creating a group map under a no-signal condition. We repeated this procedure 10,000 times and thresholded the resulting maps using the same parameters as the original analysis ($p < 0.05$). From the thresholded maps, we estimated cluster size expected by chance ($p < 0.05$). By this procedure, the threshold for cluster size at $p < 0.05$ was 24 voxels.

To test whether neural response differences between natural speech and quilted speech are accounted for by the difference in their perceived naturalness, we explored the correlation between the difference in human-rated naturalness and a measure of neural dissimilarity. Dissimilarity measures have proven to better reflect the represen-

Table 1
Breed, neurocephalic index and age for each dog participant.

Dog	Breed	Neurocephalicindex	Age(months)
Akira	Labradoodle	67.593	40
Alma	Mix	64.680	104
Barack	Golden Retriever	69.236	52
Barney	Golden Retriever	69.065	109
Bingo	Mix	77.899	37
Bodza	Golden Retriever	70.000	51
Bran	Border Collie	70.588	98
Döme	Cocker Spaniel	67.433	71
Grog	Border Collie	69.324	134
Joey	Australian Shepherd	64.681	73
Kun-kun	Border Collie	64.640	54
Maverick	Border Collie	72.277	117
Maya	Golden Retriever	70.476	97
Mini	Mix	63.062	126
Monty	Border Collie	64.005	99
Odín	Border Collie	67.307	54
Pán	Australian Shepherd	67.035	44
Sander	Golden Retriever	71.429	76

tational geometry of neural representations than classifier performance (Kriegeskorte et al., 2008; Walther et al., 2016; Diedrichsen et al., 2011; Diedrichsen and Kriegeskorte, 2017). The dissimilarities were calculated for each across-condition trial pair (e.g., NS1 vs. SS1, NS1 vs. SS2, etc.). We calculated the difference in naturalness score by obtaining the absolute of the difference between the two stimuli's average naturalness score. To assess neural dissimilarity, we created a sphere (radius = 3 voxels) around the peak in each of the three clusters that discriminated natural speech from scrambled speech; we considered all the voxels within the sphere for the analysis. The dissimilarity was defined as the Euclidean distance between the voxels' responses to the two stimuli. We repeated the procedure for all dogs and averaged the dissimilarities across trial pairs and dogs. Then we calculated the Pearson correlation between the naturalness difference and the neural dissimilarities using all trial pairs. Finally, we confirmed the correlations by conducting permutation testing. We repeated each analysis but randomly swapped each stimulus block label. We repeated the process 10,000 times and compared the correlations found in the permutations with the correlation from the analysis.

Because MVPA only discriminates between cerebral patterns but does not inform about the directionality of effects (i.e. processing preferences), to characterize speech detection in the dog brain, we narrowed our analysis to two functionally defined speech-responsive regions-of-interest (ROIs) in the primary auditory cortex. ROIs were defined by the single strongest group-level peak of an independent study's (Boros et al., 2020) speech vs. silence contrast for each hemisphere. This was a bilateral primary auditory region in the mid ectosylvian gyrus (mESG). The two ROIs were spheres (radius = 3 mm) around bilateral primary auditory cortex peaks (mid ectosylvian gyrus, coordinates 24 −18 18 and −22 −16 20). For each ROI, the average percent signal change was extracted for the individual contrast images and then analysed with a 2 × 2 × 2 mixed model ANOVA with repeated measures (speech naturalness: natural, scrambled; language familiarity: familiar, unfamiliar; hemisphere: right, left).

Finally, to address individual variability in our sample we ran two analyses. In the first analysis we correlated the neurocephalic index and the age of the participants (Table 1) with the classifier performance in the peak of each of the seven clusters from our results (three regions from speech detection and four regions from language representation results, Table 2). To calculate neurocephalic index, we used each dog's structural image and the following formula: brain width [x coordinate distance (leftmost tip of the temporal cortex; rightmost tip of the temporal cortex)] x 100) / brain length [y coordinate distance (frontalmost tip of the olfactory bulb; most posterior tip of the occipital cortex)] (Bunford et al., 2020; Hecht et al., 2019). All correlations were corrected for multiple comparisons using False Discovery Rate.

In the second analysis we performed a whole-brain representational similarity analysis (RSA) for speech detection (N-S) and for language representation (NF-NU) to explore if the representation of the stimuli in a given brain region changed in relationship with neurocephalic index or age. We first created a dissimilarity map to N-S and NF-NU, between each condition pair (N-S and NF-NU) using a searchlight approach (sphere r = 3 voxels). We then calculated the dissimilarity as the correlation distance (1 – Pearson correlation) between the response of the voxels to the N-S and NF-NU. We then created a dissimilarity map by assigning the dissimilarity value to the center of the sphere and repeated the same process for all the voxels within the brain, thus creating a dissimilarity map for every dog (Bunford et al., 2020; Connolly et al., 2012), for each condition pair. We correlated the dissimilarity values with the neurocephalic index and the age of the participants (Table 1). Finally, we confirmed the correlations by conducting permutation testing.

3. Results

The whole-brain GLM contrast All sounds (NF, NU, SF, SU) > Silence showed two clusters with the peak z-value in the bilateral mESG (coordinates 24 −20 20 and −26 −22 12; Fig. 1). These auditory cortex activity peaks were similar to those reported in a previous study using a similar contrast (Boros et al., 2020), but clusters were smaller here, perhaps due to changes in scanning parameter settings (see a direct comparison in the Supplementary Materials: Figs. S1 and S2, Table S1). We did not find any suprathreshold clusters for any of the four individual conditions > Silence contrasts.

Searchlight MVPA revealed distinct regions in the dog brain which encode features that can enable speech detection or language representation (Fig. 2). The classifier identified different activity patterns for Natural speech vs. Scrambled speech (NF + NU vs. SF + SU) in near-primary auditory regions, namely in the bilateral mid suprasylvian gyrus, and in the left caudal suprasylvian gyrus. In turn, response patterns to Natural speech in a familiar vs. Natural speech in an unfamiliar language (NF vs. NU) differed significantly in the right rostral Sylvian gyrus, the left caudal ectosylvian gyrus, the left rostral suprasylvian gyrus, and in the left precruciate gyrus (Table 2). Fig. 3 shows the multivariate intensity patterns across conditions per participant. We did not find brain patterns differentiating between Scrambled speech in a familiar language vs. Scrambled speech in an unfamiliar language (SF vs. SU), neither ones that differentiate between Familiar language vs. Unfamiliar language overall (NF + SF vs. NU + SU) in the dog auditory cortex.

The naturalness ratings for natural speech and scrambled speech stimuli were significantly different ($\text{mean}(\pm\ \text{SD})_{\text{Natural}} = 6.53(\pm\ 0.99)$, $\text{mean}(\pm\ \text{SD})_{\text{Scrambled}} = 2.12(\pm\ 1.58)$, $F_{(1,14)} = 136.05$, $p < 0.0001$). Despite scrambled speech stimuli were rated lower in naturalness than natural speech, we found variability in their naturalness ratings (range of average score per sound 1.37 to 4.06).

To test whether neural response differences in the three clusters that discriminated natural speech from scrambled speech are accounted by the difference in their perceived naturalness (perceptual account) rather than by their difference in temporal intactness (acoustic account), we calculated the Pearson correlation between the difference in human-rated naturalness and the neural dissimilarity of stimuli pairs. We found significant positive correlations in all three clusters (Fig. 4). All the significant correlations reported above were confirmed ($p < 0.05$) by a permutation test (n = 10,000, see Methods).

In the ROI analysis intended to further confirm the role of the primary auditory cortex in speech detection, a mixed model ANOVA with repeated measures revealed a significant main effect of the speech naturalness factor ($F_{1,34} = 6.1$, $p = 0.018$), evidencing a higher response to Scrambled speech than to Natural speech stimuli, regardless of language

Table 2
MVPA clusters discriminating between conditions in the dog brain.

Contrast	Regions	Voxels	x	y	z	t	p	Accuracy	Cohen's D /EM
NF+NU vs. SF+SU	R mSSG	154	19	−19	24	3.88	< 0.001	0.59	0.91 /Large
	L mSSG	52	−17	−27	20	3.29	0.002	0.55	0.78 /Medium
	L cSSG	52	−17	−29	8	3.71	< 0.001	0.59	0.87 /Large
NF vs. NU	R rSG	116	15	−5	4	3.65	< 0.001	0.62	0.86 /Large
	L PG	101	−1	9	24	3.97	< 0.001	0.63	0.94 /Large
	L cESG	86	−21	−23	12	3.85	< 0.001	0.63	0.91 /Large
	L rSSG	63	−15	−11	20	4.19	< 0.001	0.62	0.99 /Large
SF vs. SU	n.s.								
NF+SF vs. NU+SU	n.s.								

NF = Natural speech in a familiar language; NU = Natural speech in an unfamiliar language; SF = Scrambled speech in a familiar language; SU = Scrambled speech in an unfamiliar language; *R* = right; *L* =left; mSSG = mid suprasylvian gyrus; cSSG caudal suprasylvian gyrus; rSG = rostral Sylvian gyrus; PG = precruciate gyrus; cESG = caudal ectosylvian gyrus; rSSG = rostral suprasylvian gyrus; EM = Effect's magnitude; n.s. = No significant clusters.

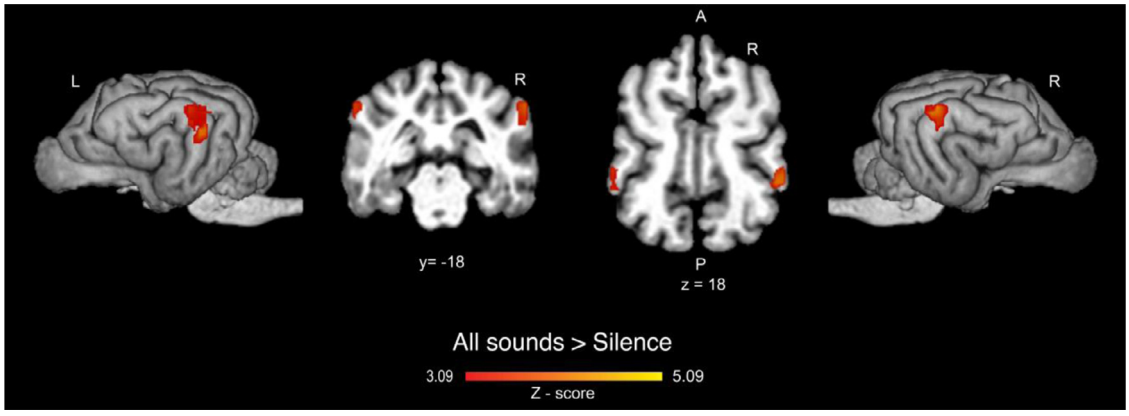

Fig. 1. GLM results of the whole-brain contrast All sounds > Silence (n = 18). Lateral, coronal, and axial views showing BOLD signal in bilateral auditory regions overlaid on a template dog brain (Czeibert, Andics, Petneházy, & Kubinyi, 2019) ($p_{unc} < 0.001$, cluster level corrected $p_{FWE} < 0.05$). *L* = Left; *R* = Right; *P* = Posterior; *A* = Anterior.

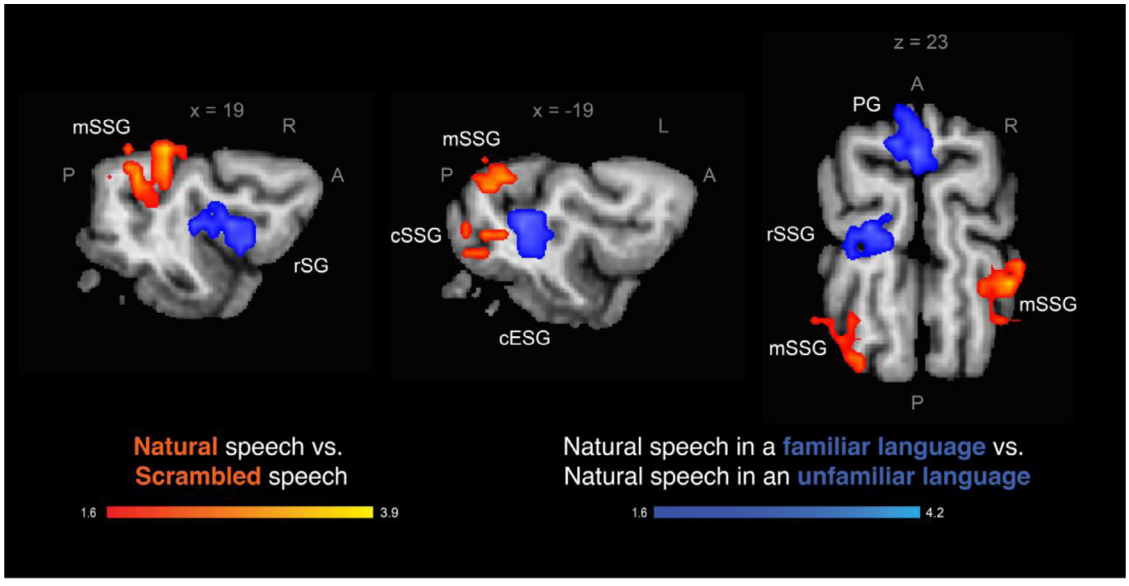

Fig. 2. MVPA results of the cerebral representation of speech and languages in the dog brain (n = 18). Clusters of searchlight analysis showing different neural representation for speech detection (red scale) and language representation (blue scale). Lateral and axial views overlaid on a template dog brain (Czeibert et al., 2019) ($p < 0.05$, cluster corrected). color bars represent the *t*-value. *L* = left; *R* = right; cESG = caudal ectosylvian gyrus; cSSG = caudal suprasylvian gyrus; mSSG = mid suprasylvian gyrus; rSSG = rostral suprasylvian gyrus; PG = precruciate gyrus; rSG = rostral Sylvian gyrus (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

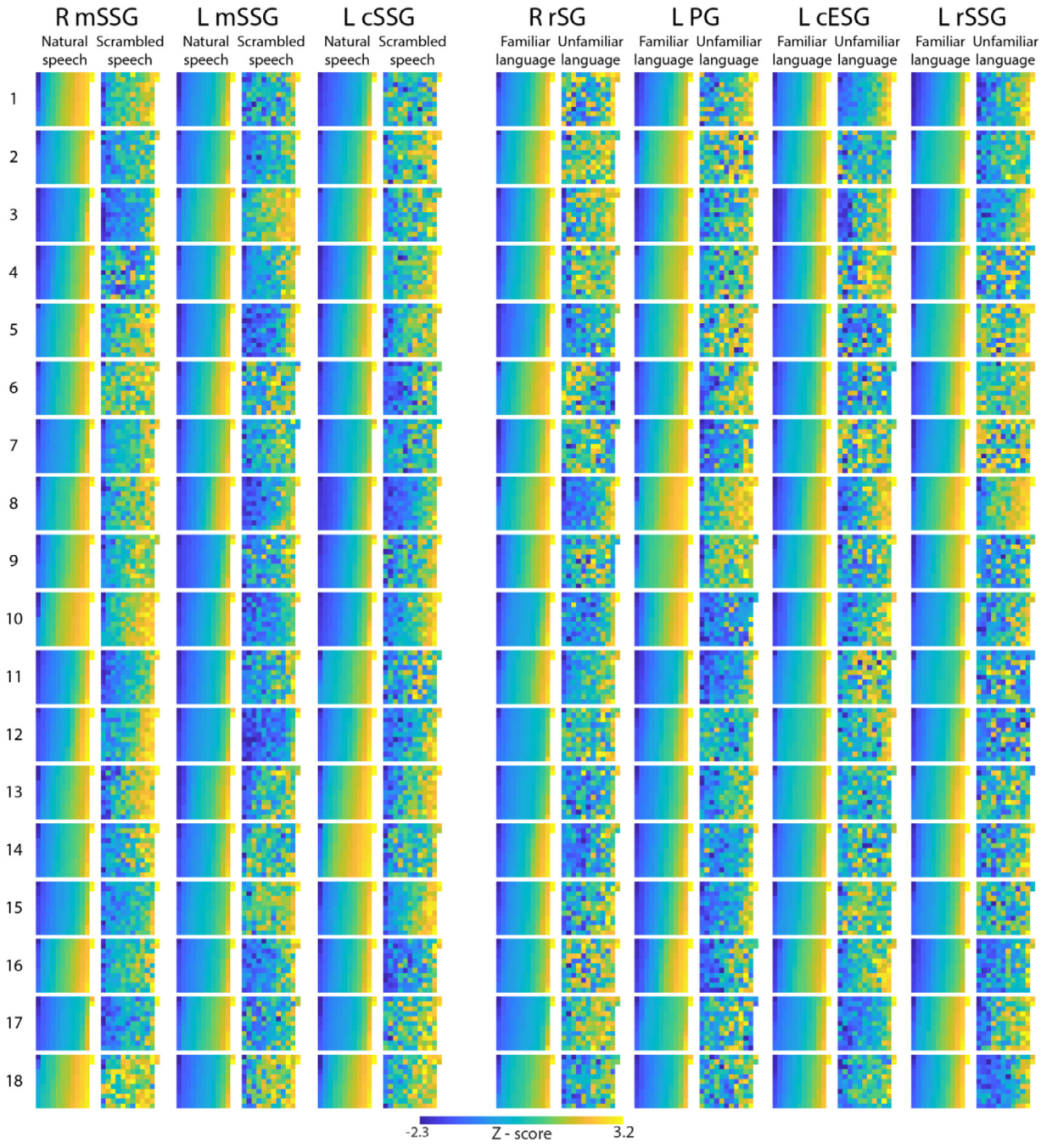

Fig. 3. Multivariate intensity patterns for each participant in the clusters identified by the searchlight analysis. Each row represents a participant (n = 18), each column represents a condition. Coloured rectangles represent voxels within the searchlight sphere (n = 123 voxels). The voxels are arranged from highest to lowest according to their response to the first condition, color-coded by their z score. The change between pattern arrangements shows the multivariate response across conditions per participant. *L* = left; *R* = right; mSSG = mid suprasylvian gyrus; cSSG caudal suprasylvian gyrus; rSG = rostral Sylvian gyrus; PG = precruciate gyrus; cESG = caudal ectosylvian gyrus; rSSG = rostral suprasylvian gyrus.

familiarity (Fig. 5). We found no significant language familiarity effect, hemispheric effect or interaction in this analysis.

The first analysis of individual differences between dogs found only a negative correlation ($r_s = -0.77$, $p = 0.005$) between the neurocephalic index and the classifier performance in the R mSSG, a near-primary auditory region (Fig. 6A). We found no correlation with the age. In the second analysis, we found no significant clusters for the correlation test between N-S dissimilarity index and either neurocephalic index or age; and neither between NF-NU dissimilarity index and neurocephalic index. But we did find two clusters with a positive correlation between NF-NU dissimilarity index and age, in the left postcruciate gyrus (PoG; 105 voxels; permutation test z-score = 2.599; $p_{cluster} = 0.012$; coordinates −13 1 24; $r_s = 0.63$, $p = 0.003$) and in the left mid suprasylvian gyrus (mSSG; 95 voxels; permutation test z-score = 3.055; $p_{cluster} = 0.013$; coordinates 15 −15 24; $r_s = 0.74$, $p = 0.0003$; Fig. 6B–D).

4. Discussion

In the present study, using fMRI MVPA, we aimed at revealing the neural representation of speech-likeness and language familiarity in dogs. We found anatomically distinct auditory cortical involvements: speech naturalness detection (differential processing of natural and

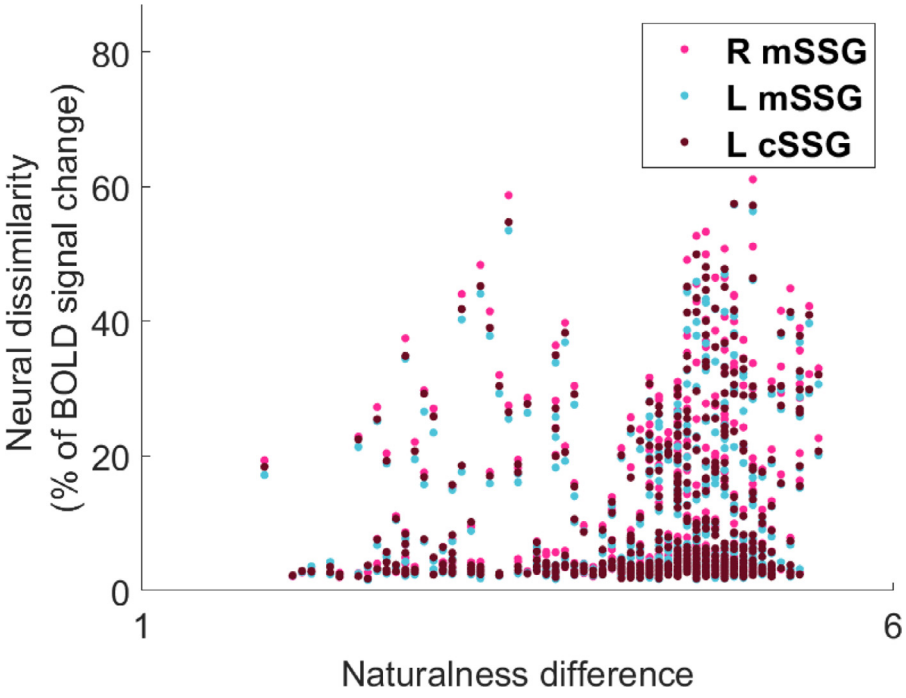

Fig. 4. Relationship between perceived naturalness difference and neural dissimilarity between natural and scrambled speech trials in the 'speech detection' clusters. Each dot represents a trial pair. The x axis shows the absolute value of the naturalness difference between the two stimuli. Naturalness was rated on a 7-point Likert scale by adult humans in a survey. The y axis represents the neural dissimilarity (percentage of BOLD signal change, calculated as the Euclidean distance between the voxels' activity patterns within the 3-voxel-radius sphere around the peak). Left: all trial pairs. Right: across-condition trial pairs only. *L* =left; *R* =right; mSSG=mid suprasylvian gyrus; cSSG=caudal suprasylvian gyrus.

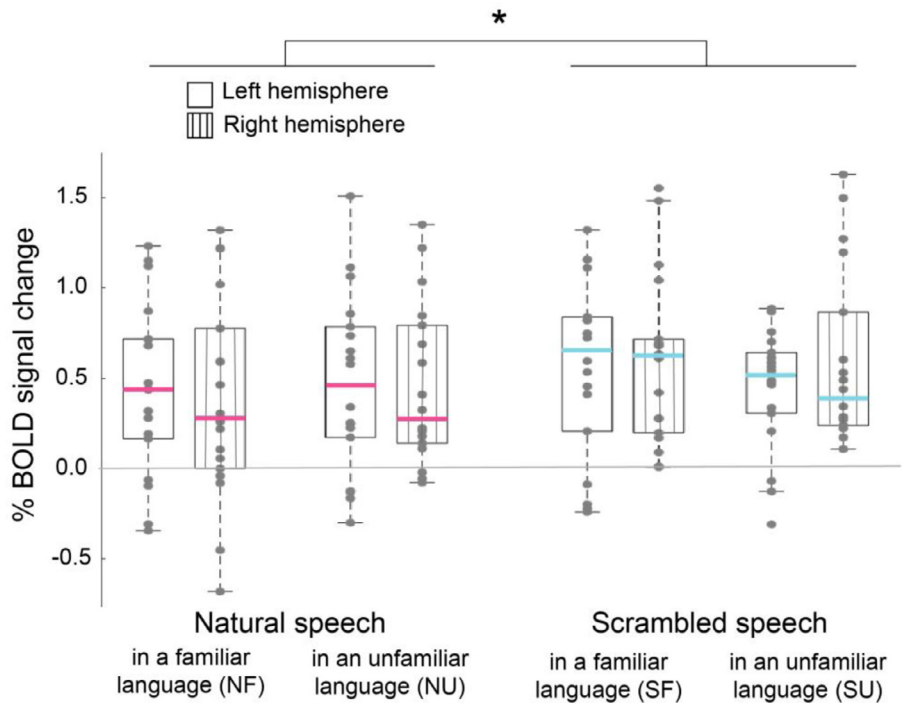

Fig. 5. ROI analysis in the dogs' bilateral primary auditory cortex ($n = 18$). Boxplot showing percentage of BOLD signal change in the bilateral mid ectosylvian gyrus in response to the four types of stimuli compared to silence. Pink lines show the median of natural speech stimuli for both hemispheres. Blue lines show the median of scrambled speech stimuli for both hemispheres. Only speech naturalness had a significant effect on the response of the primary auditory cortex, there was no significant language familiarity effect, hemispheric effect or interaction. Each gray point represents data of one dog. * $p < 0.05$.

scrambled speech) entailed bilateral near-primary auditory cortical regions, whereas language familiarity effects were found in the ventral (caudal and rostral) parts of the auditory cortex (Fig. 2). The present study provides the first evidence of distinct brain activity patterns for two languages in a non-human species.

The finding of discernable cerebral activity patterns for speech and scrambled speech stimuli in the bilateral mid suprasylvian gyri (mSSG) and the left caudal suprasylvian gyrus (cSSG) demonstrates dogs' general capacity for speech detection. These brain regions are part of the same auditory network as revealed by independent component analysis of a resting-state study (Szabó et al., 2019) and have been identified as sound-sensitive regions (Andics et al., 2016; Andics et al., 2014). It has also been proposed that the SSG is a multisensory integration cortex (Hecht et al., 2019) that also responds to familiar human social stimuli (Karl et al., 2020) and human-dog interactions (Karl et al., 2021) in the visual modality. The present results suggest that this capacity of the mSSG extends to detect naturalness in an auditory signal. On the functional level, speech detection is dependent on the precise temporal arrangement of spectro-temporal features, and the disruption of the temporal organization of speech in different time windows might have various impacts on speech recognition in humans, as evidenced by both behavioural and neuroimaging studies (Overath et al., 2015; Saberi and

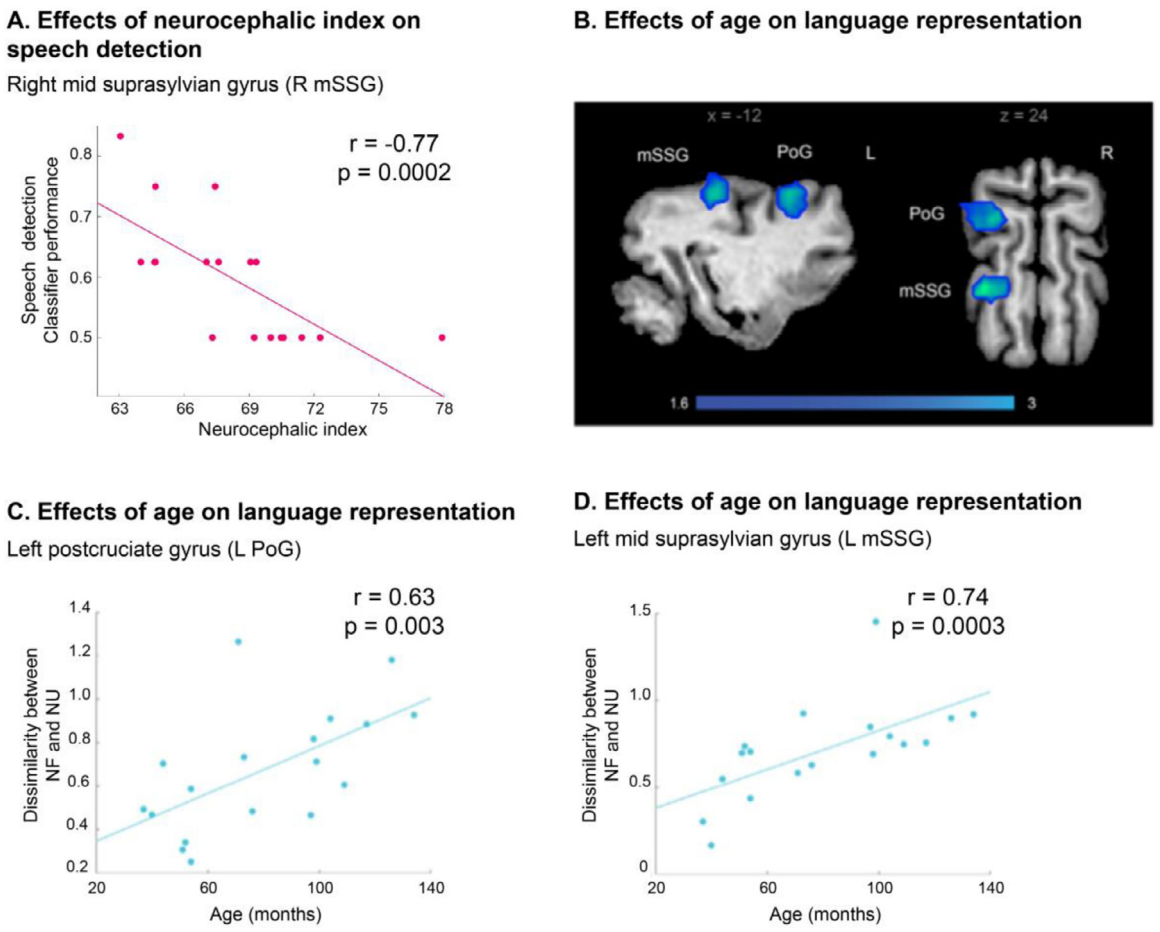

Fig. 6. Effects of neurocephalic index and age on speech detection and language representation. **A**. Effects of neurocephalic index on speech detection. Negative correlation in the R mSSG between the classifier performance in the cluster from N vs S and the neurocephalic index. **B-D**. Effects of age on language representation. **B.** Lateral and axial views showing the resulting clusters with a significant correlation between familiar and unfamiliar speech dissimilarity index and age on a dog brain template (Czeibert, Andics, Petneházy, & Kubinyi, 2019) (sphere searchlight, $r = 3$ voxels, $p < 0.05$, cluster corrected). Color bar represents the *z*-score. **C and D.** Positive correlations in the peaks of the clusters found, L PoG and L mSSG, respectively, between the dissimilarity between NF and NU [correlation distance (1 – Pearson correlation)] and the participants' age. *L* = left; *R* = right; L PoG = Postcruciate gyrus; L mSSG = mid suprasylvian gyrus; *N* = Natural speech; *S* = Scrambled speech; NF = Natural speech in a familiar language; NU = Natural speech in an unfamiliar language.

Perrott, 1999; Poeppel et al., 2008). The quilting algorithm used here to create scrambled speech has been previously (Overath et al., 2015), and also here, found successful in creating stimuli that violated the spectro-temporal features of speech, and that were also categorized as unnatural speech by human participants. In humans, pre-linguistic processing of speech-specific acoustic structure involves the Heschl's gyrus and the superior temporal cortex (e.g., Overath et al., 2015; Okada et al., 2010; Norman-Haignere et al., 2015; de Heer et al., 2017; Vouloumanos et al., 2001). Similar results were reported in macaques, where the lateral belt and parabelt showed a stronger response to speech in comparison to scrambled speech (Joly et al., 2012). Our results extend on previous human and non-human primate literature demonstrating that, in addition to supporting sound processing in general (Andics et al., 2016; Andics et al., 2014), the bilateral mSSG and left cSSG in dogs is involved in basic spectro-temporal analysis of speech. This suggests that the dog mSSG may be functionally comparable to the human superior temporal cortex and the monkey lateral belt and parabelt.

The "speech detection" classification result is evidence for dog brains' capacity to distinguish natural speech from quilted speech. Importantly, however, this may be the result of different underlying processes. (1) The underlying mechanism may be tuning to speech (this would be supported by increased responses to natural speech in directional univariate analyses (Joly et al., 2012), or a more general novelty detection mechanism (this would be supported by increased responses to unexpected stimuli, namely quilted speech in univariate analyses). (2) Natural and quilted speech may elicit differential responses because natural speech is perceived to be more natural (perceptual account), or it is temporally more intact (acoustic account). (3) This differential response to speech may be specific to speech (speech-specific process) or not (general auditory process). Below we are discussing our findings in relation to each of these issues in turn.

Tuning to speech vs. *novelty detection.* Our analyses indicate that dog auditory cortex capacity for speech detection may not reflect neural preference for speech. The ROI analyses (Fig. 5), carried out in the independently determined speech-responsive primary auditory cortex, the bilateral mid ectosylvian gyrus (mESG), suggest an opposite effect, namely stronger neural activity for scrambled than natural speech stimuli (independently from the language). The activity increase for scrambled speech in the mESG may indicate a novelty detection mechanism, or may reflect more effortful processing of unnatural acoustic stimuli, and consequently the delegation of additional neural resources (Wild et al., 2012).

Perceptual account vs. *acoustic account.* We have not tested whether quilting modulates dogs' neural response gradually, as we have used only a single quilt level (the shortest, 30 ms quilt time windows from (Overath et al., 2015), assuming that this would lead to a maximal quilt

effect). However, we tested whether dogs' brain response was modulated gradually by the (human-)perceived naturalness of individual stimuli, while keeping quilt level constant. We found that within the dog auditory regions that classified natural and scrambled speech above chance, the perceived naturalness of speech did modulate brain response patterns on a trial-by-trial basis (Fig. 4), i.e. pairwise naturalness difference between natural and scrambled speech stimuli showed significant positive correlation with the corresponding neural dissimilarity. This supports a perceptual account over an acoustic account: natural and quilted speech may have elicited differential responses because natural speech was perceived to be more natural, and not simply because it was temporally more intact.

Speech-specific process vs. *general auditory process*. We do not claim that the brain responses that differed for natural and quilted speech revealed a process specific for speech. We also do not claim the opposite, i.e. that these response patterns reflect a general auditory process. Our study was not designed to disentangle these accounts, future research will have to clarify this. The only related claim we can make here is that language familiarity did not modulate speech detection, so the process underlying speech detection in dogs is not specific to familiar language stimuli.

Together, our initial and follow-up results suggest that speech detection in dog brains may be based upon sensitivity to perceived naturalness rather than to temporal intactness of acoustic stimuli. This sensitivity is not restricted to familiar language stimuli, and does not reflect tuning to speech.

MVPA analysis indicated three clusters in which the activity pattern allows to discriminate between speech in a familiar and speech in an unfamiliar language. This result likely reflects dog brains' capacity to track auditory regularities which characterize the temporal organization of a given language, and use this implicit knowledge to build representations for a specific language even in the absence of explicit linguistic competence. In fact, statistical learning has been shown to allow non-human species to detect regularities in complex auditory patterns, including birdsongs (Chen and ten Cate, 2015) and speech (Toro and Trobalon, 2005). Note, however, that in the absence of explicit behavioural measures it remains a question whether dogs could discriminate between the different categories of stimuli by responding differentially to them outside the scanner. Nevertheless, the present fMRI study provided insights that are beyond the inference potential of behavioral studies: it revealed that separate cortical regions support speech naturalness detection and language representation in the dog brain.

Interestingly, similarly to infant studies (Nazzi et al., 1998; Nazzi et al., 2000), our results also suggest that distinct neural patterns emerge in dogs when listening to different languages, a familiar and an unfamiliar one, even if the two belong to the same rhythm class. Whether dogs could also discriminate between two unfamiliar languages remains unknown. Dog brains' ability to distinguish between languages from the same rhythm class reflects a capacity to extract auditory regularities specific to a given language. Future studies should determine the origin of this ability, and whether the familiarity with a given language is also mandatory for dogs to distinguish it from another language of the same rhythm class, similarly to human infants.

The clusters implicated in language representation were found mainly in secondary auditory cortical regions, in ventral (anterior) parts of the temporal cortex, including the left caudal ectosylvian gyrus (cESG), the left rostral suprasylvian gyrus (rSSG), and the right rostral Sylvian gyrus (rSG), and one region in the frontal cortex, the left precruciate gyrus (PG). These auditory regions have been systematically activated in dog auditory studies, showing sensitivity for fundamental frequency modulations in human and dog vocalizations in the rSG (Andics et al., 2014), and sensitivity to emotional valence in human and dog vocalizations (Andics et al., 2014) as well as to lexical meaning (Gábor et al., 2020) or voice identity (Boros et al., 2020) in the cESG. A broadly defined parietotemporal cortex, which includes the auditory regions from our results, also showed greater response for pseudowords compared to trained words, which was taken as evidence for its role in detecting novel words (Prichard et al., 2018). It is possible that in our study the auditory regions in the temporal cortex detected language novelty. The PG is considered a premotor region (Hecht et al., 2019) or supplementary motor region (Szabó et al., 2019). In humans, premotor regions are activated in speech perception tasks (Meister et al., 2007), especially in phonological judgments, but not speech comprehension, suggesting that they facilitate the perception of a sound as speech (Krieger-Redwood et al., 2013; Osnes et al., 2011), PG might be playing a similar role in dogs. Auditory processing of speech shows a hierarchical organization in humans (Okada et al., 2010; de Heer et al., 2017; Hickok, 2007). Recent findings indicate hierarchical processing of speech stimuli in the dog auditory cortex as well (Boros et al., 2020; Gábor et al., 2020). Consistent with these, the present study reveals that auditory cortical regions (including the bilateral primary auditory cortex) support the spectro-temporal analysis of speech in dogs, while secondary auditory regions and frontal regions are involved in higher level speech analysis, such as extracting language-specific auditory regularities. Language representation may involve a similar, higher level of processing as voice identity processing, recruiting only secondary and not near-primary auditory regions. We suggest that the lack of overlap between speech detection and language representation results supports our interpretation that speech naturalness detection and language representation are two separate processes.

Our analyses to address individual variability revealed a negative correlation between the neurocephalic index and the R mSSG classifier performance from the speech detection contrast (N vs. S) (Fig. 6A). This result suggests that longer-headed dogs show a processing advantage to human auditory cues. For vision, the opposite pattern has been reported, that is a processing advantage to human visual cues in shorter-headed dogs (Gácsi et al., 2009; Bognár et al., 2018). Together, this indicates modality-dependent head shape effects on communicative cue reading capacities in dogs. Besides, we found that older dogs' brain showed a greater difference between the representation of the two languages in L PoG and LmSSG (Fig. 6B–D). Age is an imperfect measure of language exposition as aging affects neural processes also in ways relatively independent of experience. However, as aging-related neural dedifferentiation (Goh, 2011) typically leads to reduced (and not increased) neural sensitivity in older individuals, including dogs processing speech (Gábor et al., 2020), we suggest that the present results (increased neural sensitivity in older individuals) can be best explained as a learning effect. This supports our interpretation that the reported language representation effects in dogs are related to learning about language regularities.

There are some limitations of this study. First, here we used a human HRF and sparse sampling acquisition which does not allow for modeling the HRF in our data. While a previous dog fMRI study using visual stimulation reported a similar HRF in the caudate nucleus to that in humans (Berns et al., 2012), recent studies suggest that the dog HRF may peak earlier, at least in case of visual stimulation. Boch et al. (Boch et al., 2021), demonstrated earlier peaking of HRF in the visual cortex, and showed that the use of a tailored dog HRF increases fMRI detection power. Another visual study also suggests an earlier peak of the dog HRF in the temporal cortex (Cuaya et al., 2016). Currently, there is no evidence that auditory HRF also peaks differently in humans and dogs, and thus most of auditory fMRI dog studies use a human HRF (e.g., Boros et al., 2020; Andics et al., 2016). Besides, here we used a slow design where an imprecise HRF does not make a considerable difference. Future studies should determine if auditory fMRI studies can benefit from a tailored dog HRF by increasing model fit and detection power. Second, due to the current technological limitations in awake dog neuroimaging (Huber and Lamm, 2017), we used a dStream Pediatric Torso coil instead of a coil designed for dogs. Our parameters aimed for the best signal-to-noise ratio (SNR) but at the cost of losing BOLD signal sensitivity with a rather short TE. Short TE, by allowing less time for dephasing and by reducing signal loss may contribute to

reducing susceptibility artifacts (Deichmann et al., 2002), a critical issue in dog fMRI because of dogs' large air cavities. Fig. S1 compares the same dogs' raw functional images from the current study and from a study with longer echo time (and larger voxel size) (Boros et al., 2020), demonstrating reduced susceptibility artifacts in the present study. Even though we used long blocks (to increase design sensitivity (Maus et al., 2010)), we cannot exclude the possibility that using a human HRF and a short TE may have reduced sensitivity. While using a human HRF and a short TE could lead to false negatives, importantly, they do not lead to false positives. Despite our short TE we were able to pick up BOLD signal in bilateral auditory regions, as showed by the contrast All sounds > Silence (Fig. 1). Third, having only one speaker per language is a further limitation of this work. Despite our effort to control for some acoustic properties in our stimuli, we cannot rule out that, to some extent, speaker-related cues played a role in the language representation results. However, the fact that we found no language effect for scrambled speech stimuli indicates that the activity difference between languages for natural speech reflected sensitivity to language-specific regularities (that were distorted by scrambling) rather than sensitivity to acoustic differences between the speakers of each language (given that scrambling kept spectral voice cues essentially intact). The present study keeps open the question whether dog brains, similarly to human brains (Belin et al., 2000; Dehaene-Lambertz et al., 2006) exhibit any special sensitivity to process speech stimuli over other natural, complex sound categories.

5. Conclusion

In conclusion, we showed that the dog brain has the capacity to detect speech naturalness and distinguish between languages, and we demonstrated that these processes are supported by different cortical regions. Speech detection in dogs may be supported by auditory cortical (including the bilateral primary auditory cortex) sensitivity to the naturalness of the acoustic signal, rather than by neural processes tuned for speech, as in humans. Longer-headed dogs' greater auditory sensitivity to speech naturalness, however, indicates breed differences in processing human auditory cues. Language representation in secondary auditory and frontal cortical regions in dogs could reflect their capacity to extract certain auditory regularities which, despite perhaps not being specific to speech, characterize the temporal organization of continuous speech in a given language. A more pronounced language representation in older dog brains suggests a role for the amount of language exposure.

Acknowledgments

We thank to our participants and their owners for their enthusiastic participation.We are grateful to Miriam Herrera-Aguilar for her help as Spanish speaker, to Tania Jasso for her fruitful comments, to Erik Pasaye for his technical comments, and to Rita Báji and Márta Gácsi for their help in data acquisition. We thank the Department of Neuroradiology, Medical Imaging Centre, Semmelweis University, Budapest.

Funding statement

This project was funded by the Hungarian Academy of Sciences via a grant to the MTA-ELTE 'Lendület' Neuroethology of Communication Research Group (grant no. LP2017–13/2017), and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 950159). LVC and RHP were also funded by the Mexican Council of Science and Technology (CONACYT, 409258 and 407590, respectively). AD was supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences, the ÚNKP- 21-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund, the Thematic Excellence Program of the Ministry for Innovation and Technology, and the National Research, Development and Innovation Office within the framework of the Thematic Excellence Program: "Community building: family and nation, tradition and innovation."

Ethical statement

All experimental procedures on dogs were approved by the National Animal Experimentation Ethics Committee (number of ethical permission: PEI/001/1490–4/2015). The sound rating survey with humans was carried out in accordance with the relevant guidelines and with the approval of the Institutional Review Board of the Institute of Biology, Eötvös Loránd University, Budapest, Hungary (reference number of ethical permission: 2019/49).

Data and code availability statement

The data (raw data, stimuli, MATLAB logs, MVPA maps, and GLM All sounds > Silence map) that support the findings of this study are available in Zenodo at [https://doi.org/10.5281/zenodo.5727656]. Any further inquiries can be addressed to the corresponding authors.

Declaration of Competing Interest

We have no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118811.

Credit authorship contribution statement

Laura V. Cuaya: Conceptualization, Writing – original draft, Methodology, Data curation, Formal analysis, Writing – review & editing. **Raúl Hernández-Pérez:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – review & editing. **Marianna Boros:** Formal analysis, Writing – original draft, Writing – review & editing. **Andrea Deme:** Writing – review & editing. **Attila Andics:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – review & editing, Writing – original draft.

References

Andics, A., Faragó, T., 2018. Voice perception across species. In: The Oxford Handbook of Voice Perception, pp. 363–392.
Andics, A., Miklósi, Á., 2018. Neural processes of vocal social perception: dog-human comparative fMRI studies. Neurosci. Biobehav. Rev. 85, 54–64.
Andics, A., Gácsi, M., Faragó, T., Kis, A., Miklósi, Á., 2014. Report voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. 24, 1–5.
Andics, A., Gábor, A., Gácsi, M., Faragó, T., Szabó, D., Miklósi, Á., 2016. Neural mechanisms for lexical processing in dogs. Science 353 (6303), 1030–1032.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.
Berns, G. S., Brooks, A. M., Spivak, M., 2012. Functional MRI in awake unrestrained dogs. PloS one 7 (5), e38027.
Boch, M., Karl, S., Sladky, R., Huber, L., Lamm, C., Wagner, I.C, 2021. Tailored haemodynamic response function increases detection power of fMRI in awake dogs (Canis familiaris). Neuroimage 224, 117414.
Bognár, Z., Iotchev, I.B., Kubinyi, E., 2018. Sex, skull length, breed, and age predict how dogs look at faces of humans and conspecifics. Anim. Cogn. 71 (4), 447–456.
Boros, M., Gábor, A., Szabó, D., Bozsik, A., Gácsi, M., Szalay, F., et al., 2020. Repetition enhancement to voice identities in the dog brain. Sci. Rep. 10, 3989.
Bosch, L., Sebastián-Gallés, N., 1997. Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. Cognition 65, 33–69.
Bunford, N., Hernández-Pérez, R., Farkas, E., Cuaya, L.V., Szabó, D., Szabó, Á., et al., 2020. Comparative brain imaging reveals analogous and divergent patterns of species- and face-sensitivity in humans and dogs. J. Neurosci. 40, 41.
Chen, J., ten Cate, C, 2015. Zebra finches can use positional and transitional cues to distinguish vocal element strings. Behav. Process. 117, 29–34.
Cohen, J., 1994. The earth is round (p <. 05). Am. Psychol. 49 (12), 997.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., et al., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618. [Internet]Available from: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5547-11.2012%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3532035&tool=pmcentrez&rendertype=abstract.

Cuaya, L.V., Hernández-Pérez, R., Concha, L., 2016. Our faces in the dog's brain: functional imaging reveals temporal cortex activation during perception of human faces. Stamatakis EA, editor. PLoS One 11 (3), e0149431 [Internet]Mar 2 [cited 2016 Mar 4]Available from:.

Czeibert, K., Andics, A., Petneházy, Ö., Kubinyi, E., 2019. A detailed canine brain label map for neuroimaging analysis. Biol. Futur. 70 (2), 112–120.

de Heer, W.A., Huth, A.G., Griffiths, T.L., Gallant, J.L., Theunissen, F.E., 2017. The hierarchical cortical organization of human speech processing. J. Neurosci. 37 (27), 6539–6557.

Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. Science (80-) 298 (5600), 2013–2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., et al., 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. Proc. Natl. Acad. Sci. 103 (38), 14240–14245.

Deichmann, R., Josephs, O., Hutton, C., Corfield, D.R., Turner, R., 2002. Compensation of susceptibility-induced BOLD sensitivity losses in echo-planar fMRI imaging. Neuroimage 15 (1), 120–135.

Diedrichsen, J., Kriegeskorte, N., 2017. Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS Comput. Biol. 13, 1–33.

Diedrichsen, J., Ridgway, G.R., Friston, K.J., Wiestler, T., 2011. Comparing the similarity and spatial structure of neural representations: a pattern-component model. Neuroimage 55 (4), 1665–1678.

Gábor, A., Gácsi, M., Szabó, D., Á, Miklósi, Kubinyi, E., Andics, A, 2020. Multilevel fMRI adaptation for spoken word processing in the awake dog brain. Sci. Rep. 10 (1), 1–11.

Gácsi, M., McGreevy, P., Kara, E., Á, Miklósi, 2009. Effects of selection for cooperation and attention in dogs. Behav. brain Funct. 5 (1), 1–8.

Goh, J.O.S., 2011. Functional dedifferentiation and altered connectivity in older adults: neural accounts of cognitive aging. Aging Dis. 2, 30–48.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7 (1), 37–53.

Hecht, E.E., Smaers, J.B., Dunn, W.D., Kent, M., Preuss, T.M., Gutman, D.A., 2019. Significant neuroanatomical variation among domestic dog breeds. J. Neurosci. 39 (39), 7748–7758.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4 (4), 131–138.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., Taga, G., 2006. The right hemisphere of sleeping infant perceives sentential prosody. Neurosci. Res. 54, 276–280.

Huber, L., Lamm, C., 2017. Understanding dog cognition by functional magnetic resonance imaging. Learn. Behav. 45 (2), 101–102. [Internet]Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.aaf3777.

International Phonetic Association IPAS, 1999. Handbook of the International Phonetic Association: A guide to the Use of the International Phonetic Alphabet. Cambridge University Press.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. NeuroImage 62 (2), 782–790.

Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012. Processing of vocalizations in humans and monkeys: a comparative fMRI study. Neuroimage 62 (3), 1376–1389. doi:10.1016/j.neuroimage.2012.05.070, [Internet].Available from:.

Jusczyk, P., Pisoni, D., Mullennix, J., 1992. Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition 43, 253–291.

Kaminski, J., Call, J., Fischer, J., 2004. Word learning in a domestic dog: evidence for 'fast mapping'. Science 304 (5677), 1682–1683. [Internet]Jun 11Available from: http://www.ncbi.nlm.nih.gov/pubmed/15192233.

Karl, S., Boch, M., Zamansky, A., van der Linden, D., Wagner, I.C., Völter, C.J., et al., 2020. Exploring the dog-human relationship by combining fMRI, eye-tracking and behavioural measures. Sci. Rep. 10 (1), 1–15.

Karl, S., Sladky, R., Lamm, C., Huber, L., 2021. Neural responses of pet dogs witnessing their caregiver's positive interactions with a conspecific: an fMRI study. Cereb. Cortex Commun. 2 (3), tgab047 tgab047.

Kohári, A., 2018. Időzítési Mintázatok a Magyar Beszédben. ELTE Eötvös Kiadó, Budapest.

Krieger-Redwood, K., Gaskell, M.G., Lindsay, S., Jefferies, E., 2013. The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. J. Cogn. Neurosci. 25 (12), 2179–2188.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. 103 (10), 3863–3868.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (November), 4.

Kuhl, P.K., 1994. Learning and representation in speech and language. Curr. Opin. Neurobiol. 4 (6), 812–822.

Lleó, C., 2003. Some interactions between word, foot, and syllable structure in the history of Spanish. In: Optimality Theory and Language Change. Springer, Dordrecht, pp. 249–283.

Martínez, S., 2011. Dialectal variations in Spanish phonology: a literature review. Echo 6 (2), 1–8.

Maus, B., van Breukelen, G.J., Goebel, R., Berger, M.P., 2010. Optimization of blocked designs in fMRI studies. Psychometrika 75 (2), 373–390.

Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., Lacoboni, M., 2007. The essential role of premotor cortex in speech perception. Curr. Biol. 17 (19), 1692–1696.

Nazzi, T., Bertoncini, J., Mehler, J., 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. J. Exp. Psychol. Hum. Percept. Perform. 24 (3), 756–766. [Internet]Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.24.3.756.

Nazzi, T., Jusczyk, P.W., Johnson, E.K., 2000. Language discrimination by english-learning 5-month-olds: effects of rhythm and familiarity. J. Mem. Lang. 43 (1), 1–9.

Nespor, M., Shukla, M., Mehler, J, van Oostendorp, M., Ewen, C.J., Hume, E., Rice, K., 2011. Stress-timed vs. Syllable-timed Languages. In: The Blackwell Companion to Phonology. John Wiley & Sons., pp. 1147–1159.

Norman-Haignere, S., Kanwisher, N.G., McDermott, J.H., 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. Neuron 88 (6), 1281–1296. doi:10.1016/j.neuron.2015.11.035, [Internet]Available from:.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I., Saberi, K., et al., 2010. Hierarchical organization of human auditory cortex : evidence from acoustic invariance in the response to intelligible speech. Cereb. Cortex 20 (10), 2486–2495.

Osnes, B., Hugdahl, K., Specht, K., 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54 (3), 2437–2445.

Overath, T., Lee, J.C., 2018. The neural processing of phonemes is shaped by linguistic analysis. Proc. Int. Symp. Audit. Audiol. Res. 6, 107–116.

Overath, T., McDermott, J.H., Zarate, J.M., Poeppel, D., 2015. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nat. Neurosci. 18 (6), 903–911.

Perrachione, T.K., Satrajit, S.G., 2013. Optimized design and analysis of sparse-sampling fMRI experiments. Front. Neurosci. 7, 55.

Pilley, J.W., Reid, A.K., 2011. Border collie comprehends object names as verbal referents. Behav. Process. 86 (2), 184–195. [Internet]Mar [cited 2013 Sep 29]Available from: http://www.ncbi.nlm.nih.gov/pubmed/21145379.

Pilley, J.W., 2013. Border collie comprehends sentences containing a prepositional object, verb, and direct object. Learn. Motiv.. [Internet]May [cited 2013 Oct 1]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S002396901300026X.

Poeppel, D., Idsardi, W.J., Van, Wassenhove V, 2008. Speech perception at the interface of neurobiology and linguistics. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1071–1086.

Pongrácz P., Miklósi Á., Csányi V., 2001. Owner's beliefs on the ability of their pet dogs to understand human verbal communication : a case of social understanding. Curr. Psychol. 20(1/2), 87-108.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59 (3), 2142–2154. doi:10.1016/j.neuroimage.2011.10.018, [Internet]Available from:.

Price, C., Thierry, G., Griffiths, T., 2005. Speech-specific auditory processing: where is it? Trends Cogn. Sci. 9 (6), 271–276.

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 62, 816–847. doi:10.1016/j.neuroimage.2012.04.062, [Internet]Available from:.

Prichard, A., Cook, P.F., Spivak, M., Chhibber, R., Berns, G.S., 2018. Awake fMRI reveals brain regions for novel word detection in dogs. Front. Neurosci. 12, 737.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., Mehler, J., 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science (80-) 288 (5464), 349–351.

Ratcliffe, V., Reby, D., 2014. Orienting asymmetries in dogs' responses to different communicatory components of human speech. Curr. Biol. 1–5. doi:10.1016/j.cub.2014.10.030, [Internet]Available from:.

Riede, T., Fitch, T., 1999. Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris). J. Exp. Biol. 202 (20), 2859–2867.

Root-gutteridge, H., Ratcliffe, V.F., Korzeniowska, A.T., Reby, D., 2019. Dogs perceive and spontaneously normalize formant-related speaker and vowel differences in human speech sounds. Biol. Lett. 15 (20190555).

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B Biol. Sci. 336, 367–373.

Saberi, K., Perrott, D.R., 1999. Cognitive restoration of reversed speech. Nature 398 (6730), 760-760.

Siptár, P., Törkenczy, M., 2000. The Phonology of Hungarian. Oxford University Press on Demand.

Stelzer, J., Chen, Y., Turner, R., 2013. NeuroImage Statistical inference and multiple testing correction in classi fi cation-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063, [Internet]Available from:.

Szabó, D., Czeibert, K., Kettinger, Á., Gácsi, M., Andics, A., Miklósi, Á., et al., 2019. Resting-state fMRI data of awake dogs (Canis familiaris) via group-level independent component analysis reveal multiple, spatially distributed resting-state networks. Sci. Rep. 9 (1), 1–25.

Toro, J.M., Trobalon, J.B., 2005b. Statistical computations over a speech stream in a rodent. Percept. Psychophys. 67 (5), 867–875.

Toro, J.M., Trobalon, J.B., Sebastián, G.N., 2003. The use of prosodic cues in language discrimination tasks by rats. Anim. Cogn. 6 (2), 131–136.

Toro, J.M., Trobalon, J.B., Sebastián, G.N., 2005a. Effects of backward speech and speaker variability in language discrimination by rats. J. Exp. Psychol. Anim. Behav. Process. 31 (1), 95–100.

Vouloumanos, A., Werker, J.F., 2007. Listening to language at birth: evidence for a bias for speech in neonates. Dev. Sci. 10 (2), 159–171.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., Liddle, P.F., 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J. Cogn. Neurosci. 13 (7), 994–1005.
Vouloumanos, A., Werker, J.F., Nelson, K., Luce, C., 2004. Tuned to the signal: the privileged status of speech for young infants. Dev. Sci. 7 (3), 270–276.
Vouloumanos, A., Hauser, M.D., Werker, J.F., Martin, A., 2010. The tuning of human neonates' preference for speech. Child Dev. 81 (2), 517–527.
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137, 188–200.
Watanabe, S., Yamamoto, E., Uozumi, M., 2006. Language discrimination by Java sparrows. Behav. Process. 73 (1), 114–116.
Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012. Effortful listening : the processing of degraded speech depends critically on attention. J. Neurosci. 32 (40), 14010–14021.
Wilson, B.M., Harris, C.R., Wixted, J.T., 2020. Science is not a signal detection problem. Proc. Natl. Acad. Sci. 117 (11), 5559–5567.
Zhao, J., Shu, H., Zhang, L., Wang, X., Gong, Q., Li, P., 2008. Cortical competition during language discrimination. Neuroimage 43 (3), 624–633. doi:10.1016/j.neuroimage.2008.07.025, [Internet]Available from:.