

REPORT

Neural mechanisms for lexical processing in dogs

A. Andics^{1,2,3,*}, A. Gábor², M. Gácsi¹, T. Faragó¹, D. Szabó^{1,2}, Á. Miklósi^{1,2}

¹MTA-ELTE Comparative Ethology Research Group, Eötvös Loránd University, Budapest, Hungary. H-1117 Budapest, Pázmány Péter sétány 1/C.

²Department of Ethology, Eötvös Loránd University, Budapest, Hungary. H-1117 Budapest, Pázmány Péter sétány 1/C.

³MR Research Centre, Semmelweis University Budapest, Hungary. H-1083 Budapest, Balassa u. 6.

*Corresponding author.

Email: attila.andics@gmail.com

Science 30 Aug 2016

Abstract

During speech processing, human listeners can separately analyze lexical and intonational cues to arrive at a unified representation of communicative content. The evolution of this capacity can be best investigated by comparative studies. Using functional magnetic resonance imaging, we explored whether and how dog brains segregate and integrate lexical and intonational information. We found a hemispheric bias for processing meaningful words, independently of intonation; an auditory brain region for distinguishing intonationally marked and unmarked words; and increased activity in primary reward regions only when both lexical and intonational information were consistent with praise. Neural mechanisms to separately analyze and integrate word meaning and intonation in dogs suggest that this capacity can evolve in the absence of language.

Full text

Various species rely on similar acoustic cues from vocalizations to infer inner states (1, 2). Human vocal comprehension also uses association of arbitrary sound sequences with meaning. Lexical items (words) are the basic building blocks of human languages but are hardly ever found in nonhuman vocal communicative systems, even though several species are capable of learning and discriminating arbitrary sound sequences (3, 4), associating vocalizations with specific meanings (4, 5), or producing human-like lexical items after extensive training (6).

Lexical processing in humans is lateralized to the left hemisphere (LH) of the brain (7). According to acoustic theories, this is caused by LH bias for rapidly changing signals (8), which is not unique to humans (9). In contrast, functional theories assume hemispheric bias for lexical representations of meaning, independent of acoustics (10). Nonhuman neural evidence for lexical processing is scarce. Hemispheric bias for broadly defined meaningfulness has been found for processing familiar, conspecific sounds (11–13), but the advantaged hemisphere varies across species (Gild-da-Costa and Hauser, 2006). The comparison of human and nonhuman neural mechanisms for processing spoken words may reveal how speech-related hemispheric asymmetries and lexical representations emerged during evolution.

Dogs present an ideal model for such investigations. Domestication has increased dogs' abilities to engage in acoustic communication with humans (14), they are more receptive than wolves to human vocal signals (15), and their behavior is more easily brought under human vocal control (16, 17). Dogs can recognize (up to ~1000) words as discriminative stimuli to retrieve different objects (18), and they process acoustic cues from human and dog vocalizations in overlapping auditory brain regions (19). Evidence for hemispheric bias in dogs listening to speech has been restricted to head-turning paradigms. A recent behavioral study found right-ear advantage when the salience of meaningful phonemic cues increased, indicating a possible LH bias (20). But the inconsistent coupling of orienting biases and lateralized acoustic processing (Fisher et al., 2009) necessitates more direct measures.

We applied functional magnetic resonance imaging to disentangle lexical versus intonational processing in awake dogs (19) (fig. S1). Verbal praises were used as stimuli because (i) human languages signal praise both lexically ("Well done!") and intonationally [higher pitch and pitch range, specific pitch contour (21)]; (ii) verbal praises are often used in dog-directed speech as social rewards; and (iii) neural evidence on reward processing mechanisms is well established (22, 23). Primary reward regions (the mesolimbic dopamine system), consisting of the ventral striatum (VS) and dopamine neurons of the ventral tegmental area and substantia nigra (VTA-SN) (24, 25), consistently respond more strongly to reward than to nonreward signals in humans, and this is also the case in dogs (26).

We manipulated lexical information (marked, praise word; unmarked, neutral word) and intonation (marked, praising tone; unmarked, neutral tone) separately. Experimental conditions involved every combination of word type and intonation (Fig. 1A): praise words in praising intonation (Pp), praise words in neutral intonation (Pn), neutral words in praising intonation (Np), and neutral words in neutral intonation (Nn). For praise words, we selected Hungarian expressions used by all test dog owners for praising. For neutral words, we used Hungarian conjunction words of similar frequency.

Typically, praise but not neutral words are spoken with praising intonation, and only Pp is used to address dogs. We assumed that Pp is meaningful to dogs whereas Nn is not, and that praise words thus contain lexical cues but neutral words do not.

We hypothesized that if dogs maintain lexical representations, neural reward responses would depend on both lexical and intonational information. In contrast, if dogs do not segregate lexical information from intonation, neural reward responses would be modulated only by intonational cues. Hemispheric bias for lexical processing would be consistent with functional theories, no lateralization for lexical cues and RH bias for intonational cues would support the acoustic account, and hemispheric bias for Pp (and perhaps Nn) relative to Pn and Np would argue for a role of familiarity.

Vocal stimuli from the dogs' female trainer were recorded. We used this single individual's voice, which was similarly well known to every dog, because speaker familiarity affects dogs' reactions to verbal utterances (27). Stimuli with praising intonation had a higher and more varying pitch, but no systematic acoustic differences were found between praise and neutral words (Fig. 1B). Non-Hungarian listeners scored intonationally but not lexically marked stimuli as more praising (28).

To assess overall lateralization, we used a bootstrapping approach with multiple cutoff thresholds (29) (Fig. 1C and table S1). Cortical responses for lexically marked (Pp and Pn) but not unmarked (Np and Nn) words were RH-lateralized. Hemispheric bias for lexically marked words was persistent across thresholds. There were no intonational or lexical-intonational effects. In a previous study (19), we found no lateralization bias for nonspeech human sounds. These findings support the functional account (10, 20) and suggest that dog brains maintain intonation-independent lexical representations of meaning. Lateralization for Pp could also be related to its higher frequency in dog-directed speech, but Pn, though rare, elicits a similar RH bias, making a familiarity-based account improbable. Dogs, unlike humans, showed a RH bias for lexical processing, supporting other reports that the advantaged hemisphere for processing meaningful sounds varies across species (Gil-da-Costa and Hauser, 2006).

To investigate regional effects, we performed random-effects tests, focusing on two sets of brain areas: (i) auditory regions responsive to speech (Fig. 1D and table S2), localized functionally, and (ii) primary reward regions, VS and VTA-SN (24, 25), defined anatomically (30). We tested for separate effects of lexical (Pp + Pn versus Np + Nn) and intonational (Pp + Np versus Pn + Nn) processing and for combined effects of lexical-intonational processing in three ways: an interaction test [(Pp versus Pn) versus (Np versus Nn)], a simple contrast or "integration" test (Pp versus Pn + Np + Nn), and a conjunction test [(Pp versus Pn)∩(Pp versus Np)].

Within auditory regions, we found intonation effects but no lexical or lexical-intonational effects. Intonation effects were only evident in the left middle ectosylvian gyrus (L mESG), with stronger responses for words with neutral intonation, independent of word meaning (Fig. 1E and table S3). We then modeled acoustic variation across stimuli, using parametric modulators. We found that L mESG activity negatively covaried with fundamental frequency (F0), paralleling findings of higher sensitivity to lower pitch in the near-primary auditory cortex in humans (31), macaques (32), and dogs (19). There was no intonation or F0 effect in the right homolog region (R mESG) but also no significant hemispheric bias (2, 28). Unlike in human studies, we found no effect of F0 change (33) (table S3). In whole-brain condition-dependent functional connectivity

analyses with auditory subregions as seeds, an intonation effect (stronger correlation for praising intonation) was only found between the L mESG (as the seed) and L caudate nucleus (CN; **Fig. 1F** and table S4). Connectivity results suggest a human-analog functional link between auditory and reward regions for processing praising intonation (**34**). The L mESG in dogs is thus involved in processing acoustic cues that are relevant for emotional intonation in both speech and nonspeech human and dog vocalizations (**19**). Analogously, emotional intonation processing in humans involves mechanisms that are not specific to speech (**35–37**).

In primary reward regions, we found combined lexical-intonational effects but no separate lexical or intonational effects. Pp elicited stronger neural responses than any other condition in the dopamine nuclei of the VTA-SN and in the CN (within the VS) (**Fig. 2A** and table S3). Dog reward regions thus respond most strongly to verbal praises when both lexical and intonational information fit. To further illustrate these findings, we calculated the maximal response for each voxel within the reward masks: Pp had the highest percentages in both the VS (Pp, Pn, Np, and Nn: 71.3, 6.2, 8.6, and 13.9%) and the VTA-SN (96.0, 0.0, 0.0, and 4.0%) (**Fig. 2B**). Similarly to humans, dogs appear to integrate lexical and intonational cues in speech to evaluate meaning in nonauditory brain regions (**38**).

We discovered three neural mechanisms of speech processing in dogs. First, there was a hemispheric bias for processing meaningful words, independently of intonation. Second, acoustic cues of affective speech intonation were processed independently of word meaning in mESG, and intonational markedness increased functional connectivity between auditory and caudate regions. Third, dogs relied on both word meaning and intonation when processing the reward value of verbal utterances. All three findings reveal functional analogies between dog and human brain mechanisms (**2, 7, 19, 34, 38**). We suggest that in a suitable ontogenetic environment, lexical representations can arise and be separated from acoustics, even in a nonprimate mammal. In dogs, specific selective forces during domestication could have also supported interspecific communicative and learning skills (**17**), but rapid evolution of speech-related hemispheric asymmetries is unlikely (**39**). Lateralized lexical processing does not appear to be a uniquely human capacity that follows from the emergence of language, but rather a more ancient function that can be exploited to link arbitrary sound sequences to meanings. What makes lexical items uniquely human is thus not the neural capacity to process them, but the invention of using them.

Fig. 1

Distinct neural patterns for lexical and intonation processing in dog brains.

(A) Experimental conditions. (B) Acoustic variation of stimuli. (C) Hemispheric lateralization test. Lateralization indices (LI) are shown across thresholds (connecting lines) and overall (horizontal bars). Positive values, right hemisphere. (D) Dog auditory regions responsive to speech (table S2). The color bar shows the range for one-sample t-test scores (12 degrees of freedom) for the speech (Pp + Pn + Np + Nn) > silence contrast. (E) Random-effects tests in auditory regions. (F) Functional connectivity tests. Random-effects test results are superimposed on an axial (z) slice [whole-brain familywise error (FWE)-corrected $P < 0.05$]; table S4]. $N = 13$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; error bars, SEM.

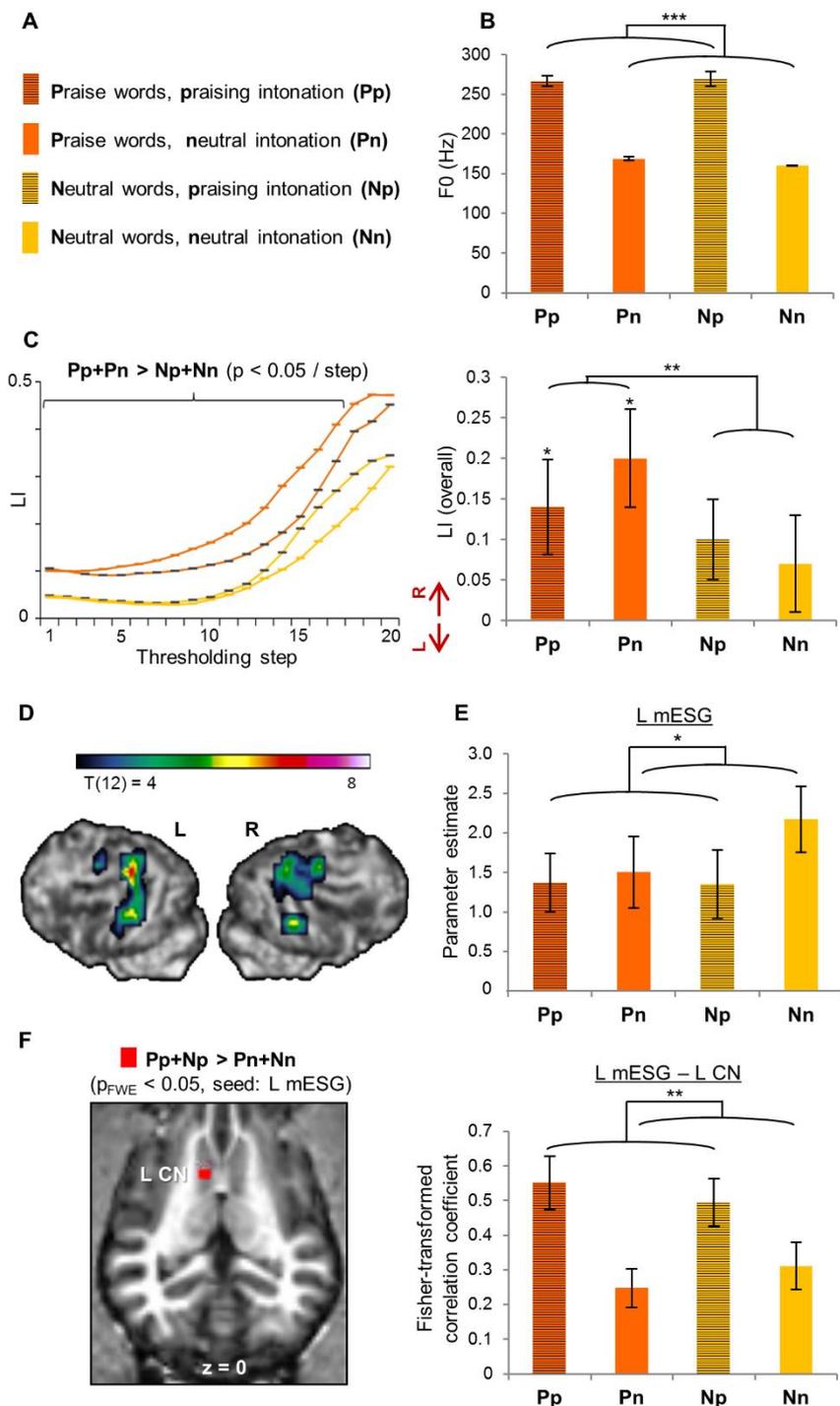
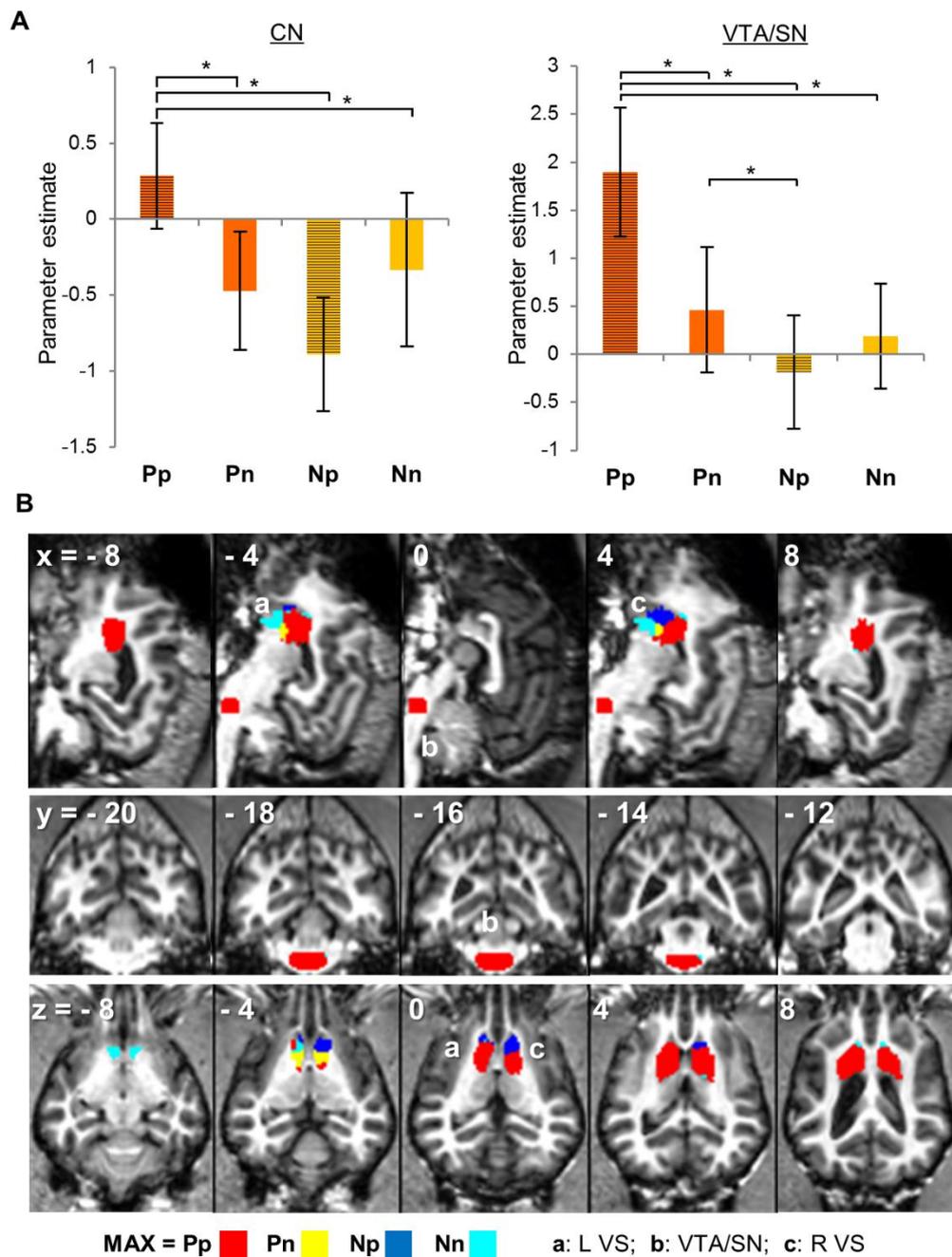


Fig. 2

Integration of lexical and intonational cues of praising in primary reward regions of dog brains.

Parameter estimates for the two activated brain areas from the integration test (small-volume FWE-corrected $P < 0.05$; table S3). Follow-up paired t tests, $*P < 0.05$. Error bars, SEM. **(B)** VS and VTA-SN masks overlaid on sagittal (x), coronal (y), and axial (z) slices, with voxels color-coded to indicate which condition elicited the maximal response.





Supplementary Materials for

Neural Mechanisms for Lexical Processing in Dogs

A. Andics, A. Gábor, M. Gácsi, T. Faragó, D. Szabó, Á. Miklósi

correspondence to: attila.andics@gmail.com

This PDF file includes:

Materials and Methods

Fig. S1

Tables S1 to S4

Materials and Methods

Participants

Thirteen family dogs, living with their owners, were tested (4 breeds: 6 border collies, 5 golden retrievers, 1 German shepherd and 1 Chinese crested; aged 1-12 years (mean = 5, SD = 2.83); 7 females and 6 males). Dog owners volunteered to participate in the training and testing process with their dogs, gave written informed consent and received no monetary compensation. Experimental procedures met the national and European guidelines for animal care and were approved by the local ethical committee (Állatkísérleti Tudományos Etikai Tanács KA-1719, Budapest, Hungary; Pest Megyei Kormányhivatal Élelmiszerlánc-Biztonsági és Állategészségügyi Igazgatósága XIV-I-001/520-4/2012, Budapest, Hungary). The training procedure, developed by MG (19) to prepare dogs for awake fMRI testing was based on conditioning and social learning. Dogs were not restricted in any way and they could leave the scanner any time.

Stimuli

Stimuli were Hungarian words, three with a praise meaning for the dogs (praise words: azaz [ˈɒzɒz] / ügyes [ˈyɟɛʃ] / jól van [ˈjoːlvɒn] for "that's it / clever / well done") and three conjunction words that presumably are not used to address dogs (neutral words: akár [ˈɒkɒːr] / olyan [ˈoːjɒn] / mégsem [ˈmeːɡʃɛm] for "as if / such / yet"). We selected praise words that, based on an open-ended questionnaire, were used by all test dog owners. The term 'neutral' is used here as a synonym for 'meaningless to the dogs' or 'non-praising and not otherwise marked'. All six words were disyllabic, with a matched consonant-vowel structure across praise and neutral words. All six words were recorded with both praising and neutral intonation. We used two tokens per word and per intonation type, adding up to 24 stimuli in total. All stimuli were spoken by a female trainer of the dogs (MG) who was also always present during the test sessions, and talked to the dogs during test preparations, but was never visible to the dogs during the actual test. To increase naturalness, stimuli were recorded as being told to a dog present in the recording room. Stimuli were digitized at a 16 bit/44.1 kHz sampling rate and were equalized for -26 dB RMS using Adobe Audition CS5.5.

Fundamental frequency parameters of the stimulus voice were compared across conditions, using a multivariate ANOVA with fixed factors lexical markedness and intonational markedness. Stimuli with a praising intonation were higher-pitched and had a greater pitch range than stimuli with a neutral intonation (praising / neutral intonation: mean(F0) = 268(±20) / 165(±6) Hz, $F_{1,20} = 289.725$, $p < 0.001$; mean(F0 range) = 277(±93) / 46(±9) Hz, $F_{1,20} = 68.264$, $p < 0.001$) but no systematic pitch differences were found between praise words and neutral words (all $F_s < 1$) (Fig. 1B). To ensure that acoustic variation in the stimuli was not specific to the stimulus voice, we analyzed recordings of the same words with both intonations from 14 other Hungarian speakers (6 males, 8 females, all familiar with dogs). Pitch parameters for these reference voices (N = 14) were compared across conditions in repeated-measures ANOVAs. For reference voices, just as for the stimulus voice, words with praising intonation were higher-pitched and had a greater pitch range than stimuli with neutral intonation (praising / neutral intonation: mean(F0) = 216(±67) / 161(±55) Hz, $F_{1,13} = 67.122$, $p < 0.001$; mean(F0 range) = 144(±71) / 37(±18) Hz, $F_{1,13} = 44.032$, $p < 0.001$) but, again, no systematic pitch differences were found between praise words and neutral words (all $F_s < 1$).

All stimuli were scored for perceived intonation on a scale of five (1: not at all praising; 5: maximally praising) by 17 foreign listeners who did not speak Hungarian and reported not to understand any of the words. Words with a praising intonation were rated as more praising (mean(words with praising intonation) = 3.826, mean(words with neutral intonation) = 1.828, $F_{1,16} = 87.741$, $p < 0.001$), while there was no difference between words with praising and

neutral meanings ($F < 1$), according to a repeated-measures ANOVA on perceived intonation scores.

Design and procedure

Four conditions were used: praise words with praising intonation (Pp), praise words with neutral intonation (Pn), neutral words with praising intonation (Np), neutral words with neutral intonation (Nn)(Fig. 1A). Stimuli were grouped into blocks of three words. Each block contained one token of each of the three words corresponding to the given condition, in a varying order. Here is an example of a stimulus block for each condition (intonation in parentheses). Pp: “azaz (praising) – ügyes (praising) – jól van (praising)”, Pn: “jól van (neutral) – azaz (neutral) – ügyes (neutral)”, Np: “mégsem (praising) – akár (praising) – olyan (praising)”, Nn: “olyan (neutral) – mégsem (neutral) – akár (neutral)”. In total, 48 stimulus blocks were created.

Stimulus blocks were presented in silent gaps between 2 s long volume acquisitions. Silent gaps were 5 s long. Stimulus onset was placed 20 ms after the beginning of the silent gap. Average word length was 642 ms (range 484 - 896 ms), and we placed 20 ms long silence periods between words within a block. Average total stimulus block length was 2.3 s. Stimulus block onset asynchrony was 7 s. Each of the 48 stimulus blocks (12 per condition) was presented exactly once per run. Twelve silent blocks were also added. Consecutive blocks were never from the same type, and all conditions were evenly distributed, but trial order was otherwise random. Every run consisted of 60 stimulus blocks, and two extra scans were added to the end. The total length of a run was approximately 7.5 mins. The experiment consisted of two runs, the order of these was counterbalanced across dogs.

Stimuli were controlled using Matlab (version 9.1) Psychophysics Toolbox 3 (40). During scanning, stimulus presentation and data acquisition were synchronized by a TTL trigger pulse. Stimuli were delivered binaurally via MRI-compatible sound-attenuating headphones that covered the ears of the dogs.

Data acquisition

MRI measurements were performed on a 3 T whole body MR unit at the MR Research Centre of the Semmelweis University Budapest. We used a Philips SENSE Flex Medium coil. We obtained EPI-BOLD fMRI time series from 29 transverse slices covering the whole brain. Spatial resolution was $3.5 \times 3.5 \times 3.5$ mm, including a 0.5 mm slice gap. A single-shot gradient-echo planar sequence was used (ascending slice order; acquisition matrix 64×64 ; TR = 7000 ms, including 2000 ms acquisition and 5000 ms silent gap; TE = 36 ms; flip angle = 90°). Both runs included 62 volumes. We also collected a standard T1-weighted three-dimensional scan using a turbo-field echo (TFE) sequence with 180 slices and $1 \times 1 \times 1$ mm spatial resolution, for anatomical reference.

Dogs were trained to lie motionless for the whole duration of the run (they were tested one run per session) without any restraints. Runs with suprathreshold motion (more than 3 mm in any direction or more than 1° rotation in any direction) were discarded. We continued test sessions with each dog until we obtained two successful functional runs, until at most three attempts with each of the two runs, and until at most four attempts in total. The average number of test sessions per dog was 2.62, the proportion of successful first / second / third attempts for a given run was 80 / 16 / 4%, respectively. Overall, twelve out of thirteen dogs completed both runs, and one dog completed only one run during these attempts.

Data analysis

We used MATLAB R2013a (<http://www.mathworks.com/products/matlab/>) and SPM8 (www.fil.ion.ucl.ac.uk/spm) for image preprocessing and analyses (41). First, the functional

EPI-BOLD images were realigned. The average of maximal movements per dog was below 1.01 mm for each translation direction, and below 0.02 degree for each rotation direction. The average of maximal movements between two consecutive volume acquisitions per dog was below 0.45 mm for each translation direction, and below 0.01 degree for each rotation direction. Individual translation movements per run, per direction are shown in Figure S1. Second, the structural image of the dog was aligned manually to the template anatomical image (a selected golden retriever, the same as in (19)). Third, the mean functional image was co-registered manually to the structural image, and the resulting transformation matrix was applied to all realigned functional images. The structural image was spatially normalized and transformed via SPM's standard nonlinear warping function with 16 iterations into a common anatomical space (the template). The same dog brain space was used as in (19), with x, y and z coordinates denoting left to right, posterior to anterior, and inferior to superior directions respectively. We then applied the resulting transformation matrix to all co-registered functional images. Finally, normalized functionals were convolved with an isotropic 3-D Gaussian kernel (FWHM = 4 mm) for spatial filtering.

The fMRI data were analyzed using a general linear model and statistical parametric mapping. We constructed three models: one with condition regressors for each run and for each block type: Pp, Pn, Np, Nn and silence; and two with parametric modulators of F0 and F0 change across all conditions (in both orders). As parametric modulators are serially orthogonalised in SPM, parametric effects were calculated for the first parametric modulator in each model. Conditions were modeled as 2.3 s long blocks. Realignment regressors for each run were also included to model potential movement artifacts. To remove low-frequency signals, we used a high-pass filter with a cycle-cutoff of 128 s. Regressors were convolved with the canonical haemodynamic response function of SPM. Single-subject fixed effect analyses were followed by lateralization tests (Table S1), small-volume corrected random effects analyses on the group level (Table S2, S3), and condition-specific seed-to-voxel functional connectivity analyses (Table S4).

To obtain a robust and specific measure of laterality which is not based on a single cutoff threshold, lateralization indices were calculated using a bootstrapping analysis approach as described in (29), and as implemented in SPM's LI-Toolbox (42). Input to this analysis were the subject-specific contrast images for each condition (compared to silence), with an exclusive midline mask of 11 mm. LIs are iteratively calculated at different thresholds. 20 thresholding intervals are used, defined as equally sized steps from 0 to the maximum value in the masked image (different per subject and per contrast, range 2.1 - 6.5; Fig. 1C). 100 bootstrapped resamples are then generated for each side, from which all 1.000 possible LI combinations are calculated for each threshold, resulting in an overall maximum of 20.000 indices. To increase the stability of the results, this approach uses trimmed mean values when analyzing LIs from each threshold, using only the mean 50% of data for each single iteration, thus providing a more representative measure of the real center of the distribution. To increase the specificity of the results, this approach takes a weighted mean of LIs from all thresholding results, based on the trimmed means obtained at each threshold. Each trimmed mean value is weighted by its respective threshold, therefore voxels with a higher contrast value get a higher weight. The process resulted in an overall weighted bootstrapped LI per subject and per contrast. These LIs across subjects and conditions were then compared in one-sample T-tests to assess condition-specific hemispheric bias, and in repeated-measures ANOVAs with lexical and intonational markedness as factors (Fig. 1C, Table S1). LIs were also calculated for frontal, temporal, parietal and occipital lobes separately, but no significant effects or tendencies were found in either test (all $p_s > 0.1$). LIs were also calculated for the human non-speech vocalizations vs. silence contrast from a previous study (19), and showed no significant hemisphere bias (mean(LI) = 0.060(\pm 0.330), $T(12) = 0.603$, $p = 0.560$).

For small-volume corrected random effects tests, brain search space was narrowed down to auditory regions and to primary reward regions. For auditory regions, we used a functionally determined mask (Fig. 1D, Table S2). This auditory mask involved the bilateral auditory cortex, extending from close-to-primary, middle ectosylvian regions to caudal and rostral ectosylvian gyrus and to the suprasylvian sulcus; two further cortical clusters in the right caudal and left middle ectosylvian gyri; and two subcortical clusters in the tectum mesencephali, near the bilateral inferior colliculi. For primary reward regions, anatomical masks were used. VTA/SN and VS were determined manually on the template brain (VTA/SN: 50 voxels, L VS: 109 voxels, R VS: 100 voxels), using anatomical atlas (30) for reference. VTA/SN is between the cerebral crura's caudo-rostral split on the basal surface of the mesencephalon and the mammillary body, directly above the surface of the interpeduncular fossa. The extent of A9 and 10 dopaminergic nuclei that constitute SN's pars compacta and the VTA is on the scale of several mm-s in each dimension. On MR anatomical images, they are indistinguishable from the interpeduncular nucleus. VS includes the nucleus accumbens, and more dorsally the caudate nucleus and the putamen (24). All regional tests are reported in Table S3. A follow-up regional test in L mESG's right homologue (R mESG) was also performed: no intonational effect or F0 effect was found, but also no significant difference across hemispheres (hemisphere \times intonation: $F_{1,12} = 1.556$, $p = 0.236$; F0: mean(R-L) = 0.003(\pm 0.011), $T(12) = 1.167$, $p = 0.266$).

For functional connectivity analyses, we applied a seed-to-voxel approach as implemented in SPM's CONN toolbox (43). Auditory mask subregions were used as seeds. Condition-specific connectivity scores (Fisher-transformed correlation coefficients) were contrasted in random effects tests. Effects are reported on the whole-brain level in Table S4.

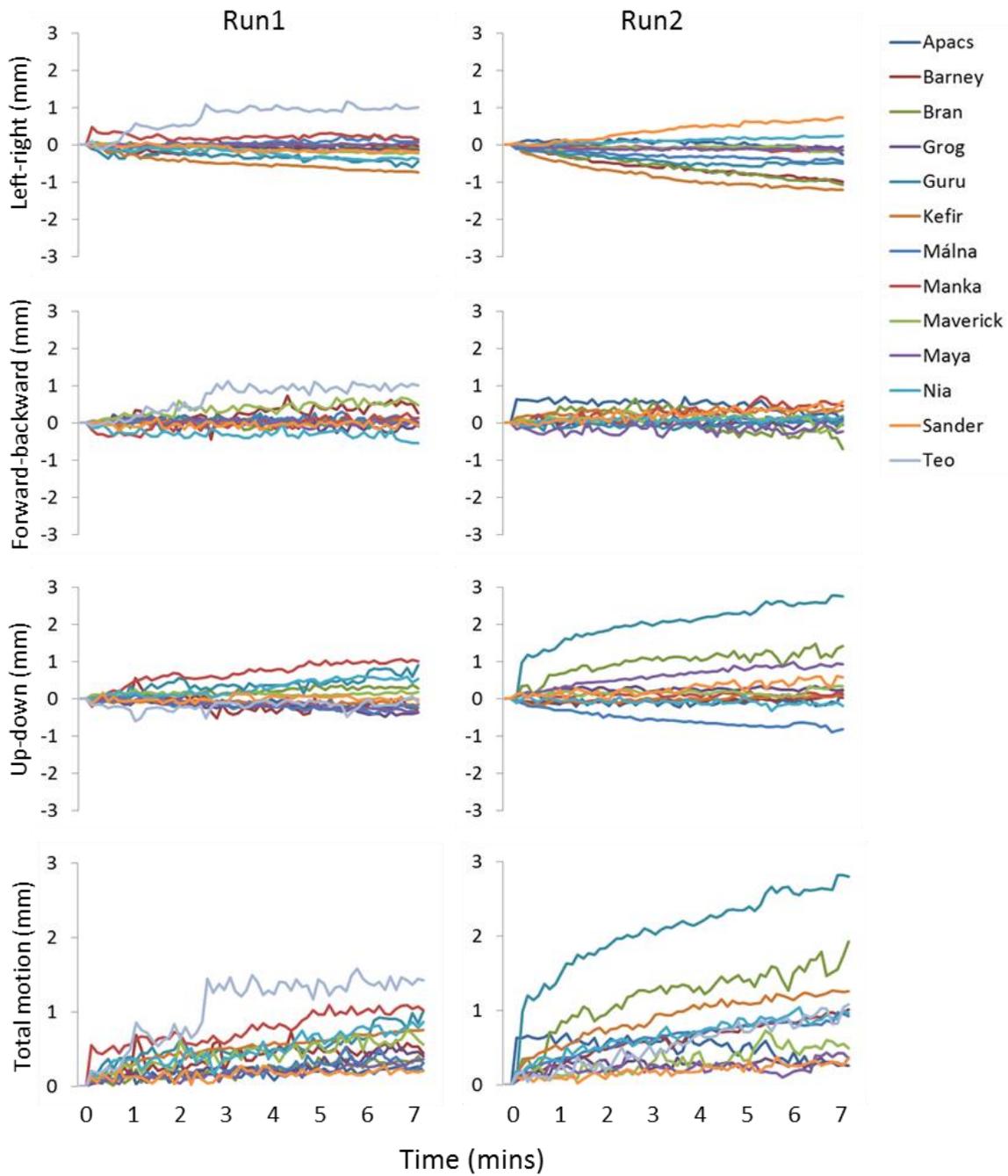


Fig. S1.

Individual dog movement parameters. Lines represent individual motion parameters for each translation direction (x axis: left-right, y axis: forward-backward, z axis: up-down) during 7 mins functional runs. Total motion refers to the Euclidean distance from the start position, calculated as the square root of $x^2 + y^2 + z^2$.

Table S1.

Overall lateralization effects across conditions and thresholding steps. Positive values denote RH bias. Effects surviving a $p < 0.05$ significance threshold are in bold. LEX: lexical, INT: intonational.

| | Pp | Pn | Np | Nn | LEX | INT | LEX × INT |
|----------------------------------|--------------|--------------|--------------|---------|--------------|-------|-----------|
| Overall LI (mean) | 0.210 | 0.284 | 0.161 | 0.116 | | | |
| | 0.026- | 0.118- | -0.017- | -0.074- | | | |
| 95% CI | 0.395 | 0.449 | 0.339 | 0.307 | | | |
| T(12) | 2.480 | 3.730 | 1.970 | 1.330 | | | |
| F(1,12) | | | | | 6.360 | 0.052 | 1.227 |
| p | 0.029 | 0.003 | 0.073 | 0.209 | 0.027 | 0.823 | 0.290 |
| p (per thresholding step) | | | | | | | |
| 1 | 0.003 | 0.001 | 0.080 | 0.060 | 0.005 | 0.824 | 0.968 |
| 2 | 0.006 | 0.001 | 0.098 | 0.071 | 0.009 | 0.954 | 0.942 |
| 3 | 0.012 | 0.001 | 0.137 | 0.110 | 0.009 | 0.916 | 0.776 |
| 4 | 0.017 | 0.001 | 0.193 | 0.160 | 0.009 | 0.824 | 0.695 |
| 5 | 0.020 | 0.001 | 0.253 | 0.209 | 0.009 | 0.722 | 0.615 |
| 6 | 0.021 | 0.001 | 0.342 | 0.275 | 0.007 | 0.748 | 0.634 |
| 7 | 0.028 | 0.000 | 0.414 | 0.318 | 0.008 | 0.670 | 0.609 |
| 8 | 0.033 | 0.000 | 0.448 | 0.402 | 0.008 | 0.639 | 0.531 |
| 9 | 0.038 | 0.000 | 0.424 | 0.421 | 0.007 | 0.616 | 0.463 |
| 10 | 0.039 | 0.001 | 0.403 | 0.373 | 0.007 | 0.587 | 0.480 |
| 11 | 0.042 | 0.001 | 0.304 | 0.324 | 0.007 | 0.557 | 0.410 |
| 12 | 0.038 | 0.001 | 0.249 | 0.276 | 0.006 | 0.541 | 0.384 |
| 13 | 0.029 | 0.001 | 0.153 | 0.201 | 0.007 | 0.522 | 0.304 |
| 14 | 0.022 | 0.001 | 0.083 | 0.180 | 0.008 | 0.551 | 0.197 |
| 15 | 0.017 | 0.001 | 0.039 | 0.158 | 0.017 | 0.733 | 0.177 |
| 16 | 0.011 | 0.002 | 0.023 | 0.120 | 0.027 | 0.927 | 0.260 |
| 17 | 0.010 | 0.002 | 0.021 | 0.102 | 0.030 | 0.985 | 0.341 |
| 18 | 0.010 | 0.002 | 0.018 | 0.088 | 0.052 | 0.934 | 0.444 |
| 19 | 0.018 | 0.004 | 0.020 | 0.078 | 0.098 | 0.992 | 0.517 |
| 20 | 0.021 | 0.018 | 0.034 | 0.085 | 0.149 | 0.985 | 0.815 |

Table S2.

Auditory regions responsive to speech. The table lists a single peak per cluster for the contrast speech [Pp+Pn+Np+Nn] > silence contrast, voxel threshold $p < 0.001$, clusters of min. 5 voxels. These voxels constituted the auditory mask.

| Brain region | x | y | z | T(12) | vol (mm ³) | p |
|---|-----|-----|-----|-------|------------------------|---------|
| Right middle suprasylvian sulcus (R mSSS) | 18 | -14 | 14 | 6.06 | 1288 | < 0.001 |
| Left middle ectosylvian sulcus (L mESS) | -28 | -10 | 8 | 7.90 | 1272 | < 0.001 |
| Right caudal ectosylvian gyrus (R cESG) | 26 | -10 | -6 | 6.49 | 232 | < 0.001 |
| Left middle ectosylvian gyrus (L mESG) | -22 | 2 | 14 | 5.30 | 64 | < 0.001 |
| Right tectum mesencephali (R TM) | 2 | -12 | -10 | 4.84 | 64 | < 0.001 |
| Left tectum mesencephali (L TM) | -4 | -8 | -12 | 4.78 | 48 | < 0.001 |

Table S3.

Random effects tests of lexical and intonational processing, and of acoustic parameters. An overall voxel threshold $p < 0.0005$ was applied. Only clusters FWE cluster-corrected for multiple comparisons across the corresponding small volume (in parentheses) are reported ($p_{\text{svc}} < 0.05$). All clusters had a single peak. n.s.: no significant clusters.

| Contrast | Brain region | x | y | z | T(12) | p | p_{svc} |
|----------------------------|--------------|-----|-----|-----|-------|----------|--------------------------|
| Lexical | | | | | | | |
| Pp+Pn > Np+Nn | n.s. | | | | | | |
| Pp+Pn < Np+Nn | n.s. | | | | | | |
| Intonational | | | | | | | |
| Pp+Np > Pn+Nn | n.s. | | | | | | |
| Pp+Np < Pn+Nn | L mESG | -20 | 0 | 18 | 4.34 | < 0.0005 | 0.041 (aud. mask) |
| Lexical and intonational | | | | | | | |
| Pp+Nn > Pn+Np | VTA/SN | 4 | -18 | -20 | 4.53 | < 0.0005 | 0.009 (VTA/SN) |
| Pp > Pn+Np+Nn | VTA/SN | -2 | -16 | -20 | 4.67 | < 0.0005 | 0.007 (VTA/SN) |
| | L CN | -2 | 10 | 6 | 5.50 | < 0.0005 | 0.022 (R VS) |
| | R CN | 6 | 10 | 4 | 4.29 | < 0.0005 | 0.022 (L VS) |
| (Pp > Pn) \cap (Pp > Np) | VTA/SN | 4 | -16 | -22 | 4.76 | < 0.0005 | 0.005 (VTA/SN) |
| | L CN | -6 | 12 | 6 | 4.64 | < 0.0005 | 0.055 (R VS) |
| | R CN | 10 | 2 | 6 | 4.48 | < 0.0005 | 0.039 (L VS) |
| F0 (parametric) | | | | | | | |
| positive | n.s. | | | | | | |
| negative | L mESG | -22 | 2 | 14 | 4.53 | < 0.0005 | 0.040 (aud. mask) |
| F0 change (parametric) | | | | | | | |
| positive | n.s. | | | | | | |
| negative | n.s. | | | | | | |

Table S4.

Seed-to-voxel functional connectivity tests. An overall voxel threshold $p < 0.0005$ was applied. Only peaks FWE corrected for the whole brain are reported ($p_{WBC} < 0.05$). n.s.: no significant clusters for any seed region.

| Contrast | Seed region | Brain region | x | y | z | T(12) | p | p_{WBC} |
|----------------------------|-------------|--------------|----|----|---|-------|----------|-----------|
| Lexical | | | | | | | | |
| Pp+Pn > Np+Nn | n.s. | | | | | | | |
| Pp+Pn < Np+Nn | n.s. | | | | | | | |
| Intonational | | | | | | | | |
| Pp+Np > Pn+Nn | L mESG | L CN | -4 | 10 | 0 | 8.99 | < 0.0005 | 0.046 |
| Pp+Np < Pn+Nn | n.s. | | | | | | | |
| Lexical and intonational | | | | | | | | |
| Pp+Nn > Pn+Np | n.s. | | | | | | | |
| Pp > Pn+Np+Nn | n.s. | | | | | | | |
| $(Pp > Pn) \cap (Pp > Np)$ | n.s. | | | | | | | |