

Published in final edited form as:

Curr Biol. 2021 December 20; 31(24): 5512–5521.e5. doi:10.1016/j.cub.2021.10.017.

Neural processes underlying statistical learning for speech segmentation in dogs

Marianna Boros^{#1,2,8,*}, Lilla Magyar^{#1,2,3}, Dávid Török^{1,2}, Anett Bozsik^{1,4}, Andrea Deme^{5,6}, Attila Andics^{1,2,*}

¹MTA-ELTE “Lendület” Neuroethology of Communication Research Group, Hungarian Academy of Sciences – Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/C, Hungary

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

³Norwegian Reading Centre for Reading Education and Research, Faculty of Arts and Education, University of Stavanger, Professor Olav Hanssens vei 10, 4036 Stavanger, Norway

⁴Department of Anatomy and Histology, University of Veterinary Medicine, 1078 Budapest, István utca 2, Hungary

⁵Department of Applied Linguistics and Phonetics, Eötvös Loránd University, 1088 Budapest, Múzeum krt. 4/A, Hungary

⁶MTA-ELTE “Lendület” Lingual Articulation Research Group, 1088 Budapest, Múzeum krt. 4/A, Hungary

These authors contributed equally to this work.

Summary

To learn words, humans extract statistical regularities from speech. Multiple species use statistical learning also to process speech, but the neural underpinnings of speech segmentation in non-humans remain largely unknown. Here, we investigated computational and neural markers of speech segmentation in dogs, a phylogenetically distant mammal that efficiently navigates humans’ social and linguistic environment. Using electroencephalography (EEG), we compared event-related responses (ERPs) for artificial words previously presented in a continuous speech stream with different distributional statistics. Results revealed an early effect (220–470 ms) of transitional probability and a late component (590–790 ms) modulated by both word frequency and transitional probability. Using fMRI, we searched for brain regions sensitive to statistical regularities in speech. Structured speech elicited lower activity in the basal ganglia, a region involved in sequence learning, and repetition enhancement in the auditory cortex. Speech

*Correspondence: marianna.cs.boros@gmail.com (M.B.), attila.andics@gmail.com (A.A.).

⁸Lead contact

Author Contributions

M.B. and L.M. conceived and designed the study, performed research, analyzed the data, and drafted the manuscript; D.T. and A.B. performed research and analyzed the data; A.D. designed the study; and A.A. conceived and designed the study and drafted the manuscript. All authors revised the manuscript.

Declaration of Interests

The authors declare no competing interests.

segmentation in dogs, similar to that of humans, involves complex computations, engaging both domain-general and modality-specific brain areas.

Introduction

The segmentation of a continuous speech signal into separate words is critical for vocabulary acquisition.^{1–3} This task is not trivial at all: stress patterns, silent pauses, and other intonational markers are not always reliable cues for word boundaries.^{4,5} A key to discover word boundaries is statistical learning: infants segment speech by computing probability distributions.^{6–8} As statistical learning is widespread in the animal kingdom (bees,⁹ chicks,¹⁰ songbirds,^{11,12} and macaques^{13,14}) but intense exposure to speech rarely results in efficient word learning in non-human species (dogs^{15,16} and apes^{17–19}), a central question is whether human computational and neural capacities for word segmentation are qualitatively different from those in other species.²⁰

Regarding computations, word segmentation in humans relies on at least two types of distributional statistics, co-occurrence frequencies and transitional probabilities, and the two differ in both complexity and efficacy.^{7,21,22} Computing co-occurrence frequencies requires simply the detection of streams of syllables occurring together, and humans already exhibit this ability at birth.^{23,24} The detection of transitional probabilities (TPs) provides information about how predictable one element is with respect to another⁷ by computing the frequency with which a transition between adjacent syllables occurs during the exposure phase relative to the frequency of that syllable occurring.²⁵ Computing TPs is more complex, but it is also more efficient for word learning, because TPs better reflect the types of dependencies that have to be learned in natural languages.⁷ In humans, infants as young as 7 months old can compute TPs from speech.²⁶

To date, the use of TPs for word segmentation in non-human animals has only been convincingly demonstrated in song-birds.^{27,28} Previous behavioral studies demonstrated statistical learning capacities for word segmentation from speech in two non-human mammals, cotton top tamarins²⁹ and rats,³⁰ but neither study provided evidence for the use of TPs in these species. However, other statistical learning tasks provide abundant evidence that macaques can perform TP computations, for example, for learning artificial grammars,^{25,31} and they rely on similar neural mechanisms as humans while doing so.^{14,32} It is unclear whether the lack of evidence for TP use for word segmentation in non-human mammals reflects their different computational capacities or the insensitivity of the experimental methods. Indeed, previous studies on statistical learning for speech segmentation in non-human mammals used performance measures, which can be inadequate to capture the learning outcomes of statistical learning tasks.³³ In contrast, methods that went beyond performance measures, such as functional magnetic resonance imaging (fMRI),³² intracranial recordings,^{13,14,27,28} or eye-tracking,^{25,34} indicated the use of TP computations in various statistical learning tasks across species. Neuroscientific measures could thus clarify the role of TP computations for word segmentation in non-human mammals.

Regarding neural substrates of statistical learning for word segmentation in humans, evidence comes from both electroencephalography (EEG) and fMRI studies. Two event-related response (ERP) components have been proposed as neural markers of word segmentation. In adults, the detection of word onset is indexed by an N100 effect,^{35–37} and recognition of segmented chunks, i.e., words, is most often implied by the presence of an N400 effect both in infants and adults.^{38–41} As for localization, statistical learning for word segmentation involves both domain-general and modality-specific brain regions in humans. Subcortical regions, including the basal ganglia^{42–47} and the hippocampus,^{44–46,48} are consistently activated in statistical learning tasks in different modalities and have long been associated with sequence or contingency learning.^{49–52} However, sound stream segmentation predominantly engages auditory cortical brain areas,^{36,40,42,43,47,53,54} suggesting the involvement of modality-specific processes. In non-human mammals, the neural substrates of statistical learning for word segmentation are unexplored.

Family dogs have recently been the subjects of several non-invasive neuroscientific studies on speech perception (fMRI^{55–58} and EEG⁵⁹). These studies identified ERP components for distinguishing known and unknown words in dogs⁵⁹ and suggested the involvement of the auditory cortex in word processing.⁵⁸ We assume that, if the computational and neural capacities to segment a speech stream are shared among mammals, then dogs, due to their domestication history⁶⁰ and language-rich experience,⁶¹ are suitable subjects for investigating this ability.

Here, using EEG and fMRI, we test the hypothesis that dogs can make use of TPs for word segmentation, and similarly to humans, they engage both domain-general and modality-specific brain areas for word segmentation. The aim of the EEG experiment was to characterize the computations used by dogs during word segmentation. To find out whether dogs make use of frequencies of co-occurrence or the computationally more complex TPs, we exposed dogs to a speech stream consisting of trisyllabic sequences (artificial words) with different distributional statistics in a ca. 8.5-min-long familiarization session. Next, we compared dogs' ERPs for these words presented in isolation. Then, in an fMRI experiment, we identified dog brain regions sensitive to statistical regularities. Using similar stimuli as during the EEG familiarization, we created statistically structured and unstructured syllable sequences and measured brain activity differences reflected in the blood-oxygen-level-dependent (BOLD) signal using fMRI, within-session repetition effects, and between-session learning effects before versus after a 16-min familiarization phase.

Results

EEG experiment

To demonstrate speech segmentation in dogs and characterize the computations supporting this capacity, we applied a similar setup as used earlier in behavioral studies with human infants⁷ and rats,³⁰ adapting it for EEG. We compared dog auditory ERPs for different conditions, as this method offers an opportunity to reveal fine temporal differences in the processing of linguistic stimuli and enables disentangling the effects of different types of statistics, even in the absence of a trained response. Using stimuli commonly used in similar studies,^{6,7,29,30} we first familiarized dogs with a speech stream in which half of

the artificially created words occurred twice as often (i.e., the co-occurrence frequency of their syllables was high, hereinafter high-frequency [HF] words) as the other half (i.e., the co-occurrence frequency of their syllables was low, hereinafter low-frequency [LF] words); thus, they differed in their word frequencies. The end of a HF word and the beginning of a consecutive HF word in the stream created partwords (PWs), which had equal word frequency to the LF words but differed in their TPs (see Table 1, Figure 1A, and STAR Methods). Next, in the test phase, HFs, LFs, and PWs were presented to dogs in isolation. We also included a non-word (NW) filler condition to keep equal occurrence frequencies for all syllables during the test. NWs consisted of the syllables of the LF words, but they were juxtaposed; thus, they never appeared in a given order during familiarization, and therefore, they were unfamiliar to dogs. Dogs' EEG was measured non-invasively following the EEG setup and analysis developed in our lab for canine studies.^{59,62}

As a first step, we selected two *a priori* time windows based on the speech segmentation indices identified by human ERP studies (the N100 and N400 components; see Sanders et al.,³⁵ Abia and Okanoya,³⁶ Abia et al.,³⁷ and Cunillera et al.⁴⁰ and STAR Methods). For each time window, we performed two separate comparisons using linear mixed models (LMMs).^{7,30} We tested (1) the effect of word frequency by contrasting ERPs for HFs and LFs (equalized for their transitional probabilities) and (2) the effect of TP by contrasting ERPs for LFs and PWs (equalized for their word frequency). Predictors in the LMM included fixed factors of TP (model 1) or word frequency (model 2) and electrode (Fz and Cz) and subject as random factor.

When testing the N100 effect (80–120 ms after word onset), we found a significant effect of electrode ($b = 1.16$; $\chi^2(5) = 5.246$; $p = 0.022$; $\eta_p^2 = 0.24$), indicating bigger deflection differences at the frontal site (Table 2). There was no significant effect of either TP or word frequency.

When testing the N400 effect (300–500 ms after word onset), we found a significant effect of TP ($\chi^2(5) = 5.237$; $p = 0.022$; $\eta_p^2 = 0.12$), indicating more positive deflection for words with high TP ($M_{(sd)} = 0.557_{(2.44)} \mu\text{V}$) than with low TP ($M_{(sd)} = -0.753_{(2.84)} \mu\text{V}$; Table 2).

Following Teinonen and colleagues,²³ we also performed a refined exploratory analysis in consecutive 50-ms time windows from 0 (word onset) to 1,000 ms that were shifted in steps of 10 ms, to reveal the precise temporal processing of the segmented words (see STAR Methods). The refined analysis identified an early (220–470 ms; $\chi^2(5) = 6.03$; $p = 0.014$; $\eta_p^2 = 0.14$; $TP_{\text{high}} M_{(sd)} = 0.242_{(2.49)} \mu\text{V}$; $TP_{\text{low}} M_{(sd)} = -1.12_{(2.76)} \mu\text{V}$) and a late (680–760 ms; $\chi^2(5) = 4.940$; $p = 0.026$; $\eta_p^2 = 0.12$; $TP_{\text{high}} M_{(sd)} = 0.157_{(1.40)} \mu\text{V}$; $TP_{\text{low}} M_{(sd)} = -0.463_{(1.39)} \mu\text{V}$) time window with a significant effect of TP (Figure 1B and Table 2). We also identified a window with a significant effect of word frequency (540–710 ms; $\chi^2(5) = 8.713$; $p = 0.003$; $\eta_p^2 = 0.20$; $WF_{\text{high}} M_{(sd)} = -0.988_{(1.32)} \mu\text{V}$; $WF_{\text{low}} M_{(sd)} = -0.065_{(1.74)} \mu\text{V}$) and another one with a significant interaction between word frequency and electrode (670–790 ms; $\chi^2(5) = 4.752$; $p = 0.029$; $\eta_p^2 = 0.12$; Figure 1B and Table 2). Post hoc analysis revealed a significant effect of word frequency only on the frontal electrode (Fz) in this time window ($t(18) = 2.59$; $p = 0.018$; $WF_{\text{high}} M_{(sd)} = -1.14_{(1.35)} \mu\text{V}$; $WF_{\text{low}} M_{(sd)} = 0.005_{(1.63)} \mu\text{V}$; Figure 1B). Model diagnostics of all LMMs showed normal distribution

of residuals, except for the model testing the late TP effect (680–760 ms), which despite transformation attempts (see STAR Methods) showed a non-normal distribution of residuals. This favors a cautious interpretation of the late TP effect (see Figure S1).

In a follow-up analysis, we also compared LF and NW conditions to confirm word frequency and/or TP effects (see also STAR Methods). This comparison was not used as a main test, because LFs had both higher word frequency and higher TP than NWs. But as the main tests showed that word frequency and TP differences lead to opposite effects (i.e., more negative deflections for higher word frequency and more positive deflections for higher TP), we assumed that more negative or positive ERPs for LF than NW would be indicative of a word frequency or a TP effect, respectively. We found more positive ERPs for the LF condition in all previously identified TP time windows (N400; 220–470 ms and 680–760 ms) and in the 670–790 ms word frequency time window as well (see Figure S2 and Table S1), all confirming the presence of TP effects.

fMRI experiment

To identify the neural structures supporting the tracking of statistical regularities in speech, another group of dogs was presented with two types of speech streams differing in their internal structure. The structured speech streams consisted of artificial three-syllable words with high within- and low between-sequence TP (i.e., with identical syllables to the ones used in the EEG experiment but with equal word frequency for all words; see Audio S1 and STAR Methods). The unstructured speech streams consisted of three-syllable-long sequences concatenated randomly; thus, they had equally low within- and between-sequence TP (Audio S1). We used similar material as used earlier in human studies.^{6,40,42,43} Dogs were measured in two fMRI sessions that were interspersed with a 16-min-long familiarization session (Figures 2A and 2B). We compared neural responses to structured versus unstructured speech streams from the baseline measurement (before familiarization) and the test measurement (after familiarization).

Acoustic stimuli elicited activity in the mesencephalon and the bilateral temporal cortices encompassing the rostral, mid, and caudal portions of the ectosylvian gyrus and a caudal portion of the Sylvian gyrus (Table 3).

An overall comparison of the conditions revealed lower activity for the structured condition as compared to the unstructured condition in a single subcortical cluster primarily located in the basal ganglia, encompassing portions of the caudate nucleus, and the thalamus (Table 3; Figures 2C and 2D). The same subcortical cluster was identified when testing the same contrast in the test measurement. We found no significant voxels for the same comparison in the baseline measurement but also no inter-action between condition and measurement in any brain region (Table 3).

The involvement of the auditory cortex in linguistic statistical learning was tested, characterizing both within-session repetition effects and between-session learning effects using the perisylvian lobe mask of the Cornell Canine Atlas.⁶³ Trial-based parameter estimates (beta values) from each repetition (1st, 2nd, and 3rd) within a session, fMRI session (baseline and test), and condition (structured and unstructured) were extracted from the

whole region of interest (ROI) of each dog individually and entered in a repeated-measures (RM) ANOVA (see STAR Methods). This analysis revealed a significant interaction between condition and repetition ($F(1,17) = 3.730$; $p = 0.034$; $\eta_p^2 = 0.180$; Tables S2 and S3).

Post hoc analysis revealed a significant effect of repetition only in the structured condition ($F(1,34) = 6.255$; $p = 0.023$; $\eta_p^2 = 0.269$), showing an activity increase over within-session repetitions (Figure 2E). No significant between-session effects were found.

Age- or head-shape-related effects

To test whether the obtained results depend on age or head shape (cephalic index), we correlated these factors (see Table S4) with individual response differences between conditions. For EEG participants, cephalic index was calculated from head length and width measured externally.⁶⁴ For fMRI participants, cephalic index was calculated from brain length and width measured on MR anatomical images (see Bunford et al.⁶⁵ for details).

For the EEG study, conditional differences were calculated for each dog by subtracting the mean amplitude for high TP words from low TP words in the 300–500 ms (N400), 220–470 ms, and 680–760 ms time windows and HF versus LF words in the 540–710 ms and 670–790 ms time windows. Amplitudes were averaged over the Fz and Cz electrodes in all time windows where there was no interaction with electrodes (N400, 220–470 ms, 680–760 ms, and 540–710 ms). For the 670–790 ms time window, where an interaction of the word frequency effect and electrode was found, conditional differences were calculated for the electrode (Fz) on which the post hoc analysis showed a significant word frequency effect. In the fMRI study, conditional differences were obtained for each dog by subtracting mean beta values for the structured condition from the unstructured condition in the significant basal ganglia cluster and by calculating the difference between the slope coefficients of the linear trendline for trial-based beta values across unstructured and structured repetition blocks. None of these comparisons yielded significant correlation effects (see also Table S5).

Discussion

The findings presented here demonstrate dog brains' capacity for statistical learning to spontaneously segment a continuous speech stream. These results extend on previous findings on speech segmentation from other non-human animals^{27,29,30} and identify for the first time the computational mechanisms and neural markers supporting this capacity in non-primate mammals.

Computational strategies

Computational mechanisms were tested in the EEG experiment, which revealed that, in dogs, statistical learning for speech segmentation is based not only on the co-occurrence frequency of syllables but also on the TPs between them. This demonstrates that humans are not the only mammal species that can apply TP-based computations to extract word boundaries from a continuous speech stream.

The timing and directionality of the ERP effects found in the EEG experiment correspond to effects reported in human literature. The early TP effect (220–470 ms after stimulus onset) is comparable to the N400 effect, the segmentation index proposed by adult human

studies,^{39,40} and coincides with the results of Teinonen and colleagues²³ in human neonates. Higher ERP amplitude for high compared to low TP stimuli is also typical in human studies^{35,39,40} and suggests that this early effect reflects violation of TP-based expectations. In a late time window (540–790 ms), we identified a word frequency effect with lower amplitudes for high-frequency than low-frequency words, in line with human literature.^{66,67} This word frequency effect overlapped with a late TP effect (680–760 ms), although the interpretation of the reported late TP effect requires caution because of the non-normal distribution of the corresponding LMM residuals. However, our follow-up comparison of low-frequency words and non-words confirmed the presence of the TP effect in all previously identified time windows, including the late TP window.

Early and late ERP effects identified in the EEG experiment can be accommodated by the neural representation theory of sequences put forward by Dehaene and colleagues.⁶⁸ According to this interpretation, early components of statistical learning are indexing cerebral mechanisms of a shallow and automatic processing of transitional probabilities, which are computed locally at syllable level, whereas late components, typically arising about 100–200 ms after early TP effects, are neuronal correlates of a higher order sequence representation of frequently co-occurring items, called chunking. It is thus possible that our late ERP effects represent a higher order recognition that acts on the word level. A similar dual pattern with early (200–300 ms) and late (650–800 ms) ERP components has been reported by Magyari and colleagues⁵⁹ using more natural stimuli in dogs presented with familiar and nonsense words. Future studies should establish whether the ERP components identified by Magyari and colleagues⁵⁹ and the present study represent functionally similar processes and whether they are characteristic indices of dogs' speech processing.

Domain generality complemented by modality specificity

Regarding brain areas underlying speech segmentation in dogs, the fMRI results revealed sensitivity to serial order information in speech stimuli in both domain-general and modality-specific regions of the dog brain. This parallels reports on statistical learning from speech in humans, which involves both domain-general subcortical and modality-specific auditory brain areas.^{42,43,45,47}

Domain-general learning was identified using whole-brain analysis, yielding a single region discriminating between conditions: the left basal ganglia exhibited overall lower activity for statistically structured than unstructured stimuli (Figures 2C and 2D). The basal ganglia cluster specifically encompassed portions of the caudate nucleus and the thalamus. These brain regions have been previously identified as containing informative voxels that discriminate between trained and untrained words in dogs.⁵⁶ The basal ganglia region is a subcortical area predominantly associated with procedural and sequence learning,^{49,51,69} but its involvement in other cognitive functions, including language, has been consistently reported in humans (for a review, see Kotz⁷⁰). Engagement of the basal ganglia in our task is in line with other evidence from language-related, rule-based learning in artificial grammar learning tasks in humans⁷¹ and non-human primates,³² as well as controlled syntactic processing tasks in basal-ganglia-lesioned patients.^{72–75}

Beyond the involvement of subcortical domain-general learning systems, the dog auditory cortex also showed sensitivity to statistical regularities in linguistic stimuli. Note, however, that the activity pattern we observed in the auditory cortex, a short-term, within-session activity increase over repeated exposures (i.e., repetition enhancement) to structured, but not to unstructured, speech, is different from the pattern found in the basal ganglia. Repetition enhancement may reflect perceptual learning⁷⁶ or novel network formation⁷⁷ for speech sound sequences. Similar repetition effects have been reported earlier in case of voice identity processing in dogs.⁵⁷ This result adds further understanding to previous findings on the distinction between known and unknown sound sequence processing in the dog auditory cortex.⁵⁸

Whether the obtained sensitivity to statistical structure in the basal ganglia and the auditory cortex is specific to speech should be established by future research. Here, we did not employ control stimuli targeted specifically for speech likeness (e.g., distorted speech). Statistical learning is often claimed to be domain-general,⁷⁸ although some studies suggest that its efficiency depends on the ecological validity of the input type for a given species.²⁰ An advantage for statistical learning of language over other inputs has been demonstrated in humans^{79,80} and should be systematically tested in dogs as well.

Statistical learning and individual differences

Whereas the present study showed that dogs can apply complex computations to segment a speech stream, the origin of this ability is less clear. The capacity for word segmentation could either reflect a general mammalian capacity that was previously undetected in other non-primate mammals or could be the result of effects of exposure or domestication. Although the design of the present study does not allow to disentangle these accounts, we did test whether individual differences in word segmentation capacity are influenced by dogs' age (a potential though clearly imperfect proxy for exposure; cf. Salthouse⁸¹) or head shape (an index of breeding that has been related to communicative cue-

reading abilities; cf. Bognár et al.^{64,82} and Gácsi et al.⁸³), but we found no significant correlations with either factor. In future studies, domestication-related effects could be tested by comparing modern versus ancient breeds (differing in genetic ancestry⁸⁴) or similarly socialized wolves versus dogs^{85–88} with comparable speech exposure.

Despite similar word-segmenting abilities and sharing a linguistic environment, dogs have limited word-learning capacities.⁸⁹ Even though there is growing evidence for some similar mechanisms between dogs and humans for word processing,^{55,58,59} it is extremely rare for dogs to develop a vocabulary exceeding a dozen words.^{90–93} Our results suggest that the apparent inability of most dogs to acquire a sizeable vocabulary is not explained by computational or mechanistic limitations to use statistical learning to segment speech. A previous study found that the ERPs of dogs, despite their auditory ability to discriminate speech sounds, did not differ between processing known and phonetically similar nonsense words, suggesting that dogs may allocate their auditory attention differently than humans.⁵⁹ Another difference in human and dog ability for language learning might stem from capacities for word representations. Here, we found no between-session effects for structured stimuli in the auditory cortex. This may indicate that, in dogs, long-term word

representations do not develop as easily as in humans (cf. Cunillera et al.⁴⁰ and McNealy et al.^{42,43}), although this topic warrants further investigations. So the origins of dogs' limited language capacity may not lie in differences in statistical learning abilities but in other key competences necessary for language acquisition.

Conclusions

This study demonstrates for the first time the use of statistical learning for linguistic stimuli in dogs. Our results show that spontaneous speech segmentation occurs in dogs already after a short exposure to a continuous speech stream (ca. 8.5 min) and engages similar computational mechanisms as earlier demonstrated in human infants. The capacity for statistical learning from speech in dogs is supported by a sensitivity to serial order information in the basal ganglia and the bilateral auditory cortex. By demonstrating human-analog computational and neural markers for speech segmentation in dogs, these results extend on previous findings on speech segmentation in non-human species and provide evidence for a potentially conserved statistical learning mechanism, which can be adapted for detecting words in human speech.

Star Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
EEG data, fMRI data and scripts	This paper	[figshare]:[10.6084/m9.figshare.16698454]
Experimental models: organisms/strains		
Domestic Dog	Family dogs	https://familydogproject.elte.hu/about-us/our-research/
Software and algorithms		
Profivox Hungarian TTS software	Kiss et al. ⁹⁴ and Olaszy et al. ⁹⁵	http://smartlab.tmit.bme.hu/index-en
MATLAB 2014b	The MathWorks	https://uk.mathworks.com/
MATLAB 2018a	The MathWorks	https://uk.mathworks.com/
Psychtoolbox 3	Kleiner et al. ⁹⁶	http://psychtoolbox.org/
FieldTrip	Oostenveld et al. ⁹⁷	https://www.fieldtriptoolbox.org/
R version 1.3.1093	R Development Core Team ⁹⁸	https://www.r-project.org/
<i>lme</i> (R package)	Pinheiro et al. ⁹⁹	https://cran.r-project.org/web/packages/nlme/index.html
<i>anova</i> (R package)	Fox and Weisberg ¹⁰⁰	https://cran.r-project.org/web/packages/car/
SPM12	Wellcome Department of Imaging Neuroscience, University College London	https://www.fil.ion.ucl.ac.uk/spm
<i>ezANOVA</i> (R package)	Lawrence ¹⁰¹	https://cran.r-project.org/web/packages/ez/index.html
Other		
NuAmps amplifier	Compumedics Neuroscan	https://compumedicsneuroscan.com/nuamps-2/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
EC2 Grass Electrode Cream	Grass Technologies, USA	https://neuro.natus.com/products-services/neurodiagnostic-supplies/emg-supplies
Philips Achieva 3 T	Philips Medical Systems	https://www.usa.philips.com/healthcare/product/HC889204/diamondselectachieva30txrefurbishedmrs scanner
MRI-compatible sound-attenuating headphones	MR Confon	http://www.mr-confon.de/en/products/headphones.html
Bluetooth speaker	JBL Go	https://eu.jbl.com/

Resource Availability

Lead contact—Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Marianna Boros (marianna.cs.boros@gmail.com). There is no restriction for distribution of materials.

Materials availability—This study did not generate new unique reagents.

Experimental Model And Subject Details

EEG experiment—The EEG data of 19 healthy companion dogs were analyzed (11 males, 1-10 years, mean age 5.4 years, breeds: 1 Akita Inu, 1 Australian Shepherd, 1 Cane Corso, 1 Dachshund, 4 Golden retrievers, 3 Great Pyrenees, 2 Hungarian Greyhounds, 1 Irish Terrier, 1 Labradoodle, 1 Small Munsterlander and 3 mixed breed dogs, see also Table S4). The owners of dogs volunteered to take part in the project without any monetary compensation and gave written informed consent. Experimental procedures met the national and European guidelines for animal care and were approved by the local ethical committee (Pest Megyei Kormányhivatal Élelmiszerlánc-Biztonsági Növényés Talajvédelmi Főosztály, PE/EA/853-2/2016, Budapest, Hungary). We initially recruited 51 dogs, from these we excluded 22 dogs who didn't preclude electrode placement or showed signs of unrest during testing (see Magyari et al.⁵⁹). Additional 10 dogs were excluded because of high number of movement-artifacts or noisy electrode during the automatic data cleaning procedure (see later). The ratio of dogs invited for the study and dogs whose data was included in the data analysis (37%) is similar to the inclusion rate of ERP studies with human infants (e.g., Forgács et al.¹⁰²) and a previous study with dogs.⁵⁹

fMRI experiment—The fMRI data of 18 healthy companion dogs were analyzed (7 males, 2-12 years, mean age 5.4 years, breeds: 1 Australian Shepherd, 3 Border Collies, 2 English Cocker Spaniels, 4 Golden Retrievers, 2 Labradoodles, 1 Springer Spaniel, 1 Tervueren and 4 mixed breed dogs, see also Table S4). This sample was independent from the sample used in the EEG study. The owners of dogs volunteered to take part without any monetary compensation and gave written informed consent. Experimental procedures met the national and

European guidelines for animal care and were approved by the local ethical committee (Pest Megyei Kormányhivatal Élelmiszer-lánc-Biztonsági, Növény- és Talajvédelmi Főosztály, PEI/001/1490-4/2015, Budapest, Hungary). The dogs were previously trained to lie

motionless in the scanner for more than 8 minutes. Details of the training procedure are described elsewhere.¹⁰³ Initially 21 dogs participated in the study (8 males, 2-12 years, mean age 5.4 years). Two dogs were excluded from the final analysis due to technical failure during testing, and another one because of the inability to preclude the baseline scanning.

Method Details

Stimuli and procedure

EEG experiment: The experiment consisted of two parts – familiarization and test – administered directly one after the other to avoid the removal of electrodes between the two parts (Figure 1A).

For the familiarization phase two artificial speech streams (language 1 and language 2) were created following Aslin and colleagues.⁷ In each language high-frequency (HF) words appeared 240 times (2 words per language) and low-frequency (LF) words appeared 120 times (2 words per language) amounting to 720 words in a 8:24 mins (language 1) and 8:41 mins (language 2) of stream exposure. Words were first concatenated as a text and then synthesized with co-articulation into an audio block using the Profivox Hungarian TTS software.^{94,95} The text was read by a female voice, with mean syllable duration in language 1 = 233.28 ms (SD = 20.37) and mean duration in language 2 = 241.13 ms, (SD = 18.77). Pitch was kept constant throughout the speech stream (182.75 Hz for language 1 and 179.23 Hz for language 2), without pitch rise or fall, creating a flat speech stream containing no acoustic markers and thus no other cues to word boundaries than the different distributions of TPs between the syllables.^{6,7,42,43} Each dog listened to one of the languages, with half of the dogs exposed to language 1 (10 dogs) and the other half to language 2 (9 dogs).

For the test phase, three-syllable long test stimuli were synthesized in isolation using the same parameters as for the speech stream.^{94,95} Four types of stimuli were created (see Table 1 and Figure 1A). 1) HF words were the same as the HF words spoken 240 times during the familiarization phase – they had high frequency of co-occurrence as well as high transitional probabilities between the syllables. 2) LF words were the same as the LF words that appeared 120 times during the familiarization – the frequency of co-occurrence between their syllables was lower than in case of HF words, but the transitional probability between them was as high as for the HF words. 3) Partwords (PW) were formed from the word boundaries of adjacent high-frequency words, i.e., they were formed using the last syllable of the first high-frequency word and the first two syllables of the second high-frequency word. The frequency of co-occurrence of their syllables was the same as for the LF words, while the TP between their syllables was lower than for HF and LF. 4) Finally, we included a non-word (NW) filler condition to keep equal occurrence frequencies for all syllables and thus to control for syllable frequency-based effects during the test. NWs consisted of the syllables of the low-frequency words, but they were juxtaposed, thus they never appeared in a given order during familiarization, and therefore they were unknown to dogs. Such a stimulus composition allowed us to test two different statistics dogs might use: frequency of co-occurrence between the syllables (here-inafter word-frequency) and the transitional probabilities between the syllables.

As it is apparent from Table 1, the two artificial languages contained the same words but in different conditions. The word that was HF in one language became LF in the other language, and vice versa. This juxtaposition made it possible to control for the effect of words on conditions. When ERPs for the LF and HF conditions are compared, approximately the same words participate in both conditions.

EEG recording took place at the laboratory of the Ethology Department of the Eötvös Loránd University in Budapest. For a general habituation protocol of companion dogs to the laboratory setting and the electrode application see also Magyari et al.⁵⁹ Dogs were first allowed to freely explore the room, and once they seemed calm, they were seated with their owner on a mat. Owners were asked to comfort their dogs with praise and strokes so that they lie down and relax, only then we proceeded to the application of electrodes. When a dog stood up during the experiment, the owner was asked to try to put the dog back to lying position. If the dog wanted to leave from the mat, the experiment was aborted.

Auditory stimuli were played from two loudspeakers placed on the ground in the right and left front of the dog in 1.5 m distance. Dogs laid down on a mat with their heads between the two speakers. During EEG recording, owners were wearing sound-proof headphones, owners and dogs were not facing each other and dogs were video recorded.

During the familiarization phase, dogs listened to the continuous artificial language stream without pauses separating each word or syllable. In the test phase each word was presented 40 times (80 trials per condition, 320 trials in total) in a pseudorandom order, with a criterion that two identical words could not follow each other directly. Stimulus onset asynchrony was set between 2500-3000 ms. The test phase took approx. 12 minutes. A short 10 s pause between the exposure and test phase was inserted.

Loudness was calibrated to 65 +/- 10 dB. The auditory stimuli and the triggers to the EEG amplifiers were presented by Psychotool-box⁹⁶ in MATLAB 2014b (The Mathworks, Inc., Natick, Massachusetts).

fMRI experiment: Statistical learning was tested by comparing brain activation to continuous speech streams with different distributional statistics. Both speech streams were built up from 12 syllables (two different syllable sets: syllable set A, containing the same syllables as language 1 and 2 in the EEG experiments, and syllable set B), but in the Structured condition they formed four artificial three-syllable words (see Audio S1), while in the Unstructured condition their order was random (see Audio S1). That is, in the Structured condition within-word TP was always higher than between word TP, while in the Unstructured condition any subsequent three syllables had similar TPs, always lower than the TPs of the Structured condition (Figures 2A and 2B).

For each condition we created text blocks containing 10 repetitions of the two syllable sets (120 syllables in each block). These text blocks were then synthesized into an audio block using the Profivox Hungarian TTS software.^{94,95} The text was read by a female voice, and was synthesized with an input of same mean syllable duration (267 ms) and pitch (190 Hz), without pitch rise or fall, creating a flat speech stream, containing no acoustic markers. To

avoid signal discontinuities in segment-to-segment transitions, the text was synthesized with a co-articulation option, which renders the speech stream more natural and thus can better sustain dogs' attention in the scanner. This resulted in slightly varying syllable durations (mean duration in syllable set A = 236.09 ms, SD = 39.08, mean duration in syllable set B = 254.52 ms, SD = 20.24). Syllable sets were counterbalanced between dogs, and thus syllable duration effects were equaled out for the whole experiment (Tables S6 and S7). Therefore, any condition differences found between the Structured and Unstructured stimuli could not be explained by systematic acoustic differences between syllable set A and syllable set B. There were no other cues to word boundaries in the speech streams than the different distributions of TPs between the syllables.^{40,42,43}

These audio blocks were then presented to the dogs in three consecutive sessions (Figure 2A). Session one was a baseline fMRI measurement in which dogs were presented with six audio blocks (three blocks of the Structured and three of the Unstructured condition). Their order was random with the criterion that all three blocks of the same condition were never consecutive. The audio blocks were interleaved with alternating 16 s or 24 s silence blocks. Next, in session two, the familiarization phase, the dog was placed with its owner in a separate room, where the dog could move freely and 16 new audio blocks from each condition (32 in total) were played back using a portable Bluetooth speaker (JBL Go). The owner was seated in the middle of the room, s/he was asked to wear sound attenuating headphones in order to prevent giving the dog any stimulus related cues and was instructed to restrict contact with the dog to minimum. The audio blocks were presented pseudo-randomly in the sense that more than two blocks from the same condition could not follow each other directly. No silence blocks were inserted between the audio blocks; thus all audio blocks were synthesized as two 8-minute-long speech streams (enabling a break if needed) using the same parameters as described above. Upon finishing with the familiarization phase, dogs returned to the scanner and were measured by fMRI in the third session, the test measurement. In the test measurement the dogs were subjected to the same procedure as in the baseline measurement, but using different audio blocks with different randomizations.

All three sessions were administered on the same day, followed each other directly, but allowed the dogs and their owners to keep a short break if needed. This resulted on average 20 minutes elapsing between the first and second fMRI measurement.

In order to control for the unlikely effect of stimulus saliency in one of the stimulus sets, half of the dogs (N = 9) was presented with the Structured condition built from syllable set A and with the Unstructured condition built from syllable set B, and the other half (N = 9) the other way around (see Table S6). Subsequent analysis confirmed, that any variation found in our stimuli were counterbalanced across conditions and dogs, and therefore could not affect the obtained results (see Table S7).

Data Acquisition

EEG experiment

Electrophysiological recording: Electrode-placement followed a canine EEG setup developed and validated in our lab, which has been used successfully to study word processing in dogs.⁵⁹ Surface attached scalp electrodes (gold-coated Ag|AgCl) were fixed with EC2 Grass Electrode Cream (Grass Technologies, USA) on the dogs' head and placed over the anteroposterior midline of the skull (Fz, Cz, Pz), and on the zygomatic arch (*os zygomaticum*), next to the left (F7, electrooculography, EOG) and right eyes (F8, EOG). The ground electrode was placed on the left musculus temporalis. Fz, Cz and EOG derivations were referred to Pz.^{59,62,104} The reference electrode (Pz), Cz and Fz were placed on the dog's head at the anteroposterior midline above the bone to decrease the chance for artifacts from muscle movements. Pz was placed posteriorly to the active electrodes on a head-surface above the back part of the external sagittal crest (*crista sagittalis externa*) at the occipital bulge of dogs where the skull is usually the thickest and under which either no brain or only the cerebellum is located depending on the shape of the skull,¹⁰⁵ see more explanation of the electrode placement in Magyari et al.⁵⁹ Impedances for the EEG electrodes were kept at least below 15 k Ω , but mostly below 5 k Ω . The EEG was amplified by 40-channels NuAmps amplifier (Compumedics Neuroscan) with applying DC-recording and it was digitized at 1000 Hz sampling-rate.

EEG artifact-rejection and analysis—EEG artifact rejection and analysis was done using the FieldTrip software package⁹⁷ in MATLAB R2018a (The Mathworks, Inc., Natick, Massachusetts).

EEG data was digitally filtered offline with a 0.3 Hz high-pass and 40-Hz low-pass filter and segmented between 200 ms pre-stimulus onset and 1000 ms after stimulus onset. Each segment was baselined between 200 ms and 0 ms (stimulus onset). Following a recent dog ERP paper from the same lab⁵⁹ showing no major differences between a multi-step data cleaning and an automatic amplitude-based rejection method, we decided to rely on the automatic artifact rejection algorithm. For this reason, automatic artifact rejection was applied after baselining by removing each segment with amplitudes exceeding ± 95 mV or with minimum and maximum values exceeding a 115 mV at any of the electrodes (Fz, Cz, F7 (left EOG) and F8 (right EOG)) in 100 ms long sliding windows. In case of one dog one of the EOG channels (F8) was excluded because the electrode fell off during measurement. EOG channels were only used for the purpose of data cleaning and were not used in further analysis. Segments were averaged separately for each condition, participant and electrodes (Fz, Cz). After the automatic artifact rejection an average of 54.7 trials remained in the HF condition (min: 38, max: 77), 56.1 trials in the LF condition (min: 37, max: 76), 54.6 trials in the PW condition (min: 31, max: 76) and 55.7 trials in the NW condition (min: 38, max: 76).

fMRI experiment: Scanning was carried out at the MR Research Centre of the Semmelweis University Budapest on a Philips Achieva 3T whole body MR unit (Philips Medical Systems, Best, the Netherlands), using an eight-channel dStream Pediatric Torso coil. For the placement of the dogs in the scanner please see Andics et al.¹⁰³ Stimuli were presented

at a 68 dB volume level using MATLAB (version 7.9) Psychophysics Toolbox 3⁹⁶ and delivered binaurally through MRI-compatible sound-attenuating headphones (MR Confon, Magdeburg, Germany). In order to minimize noise during stimulus presentation a so-called sparse sampling procedure was applied (see Andics et al.^{55,103} for a similar protocol) in which an 8 s repetition time (TR) consisted of 1.7 s volume acquisition and 6.3 s silent gap. Using such a scanning protocol one audio stream (ca. 32 s long) spanned over 4 silent gaps and 4 volume acquisitions, thus a concurrent ending of a silent gap and an audio block was minimized in order to prevent giving additional cues to word boundaries. EPI-BOLD fMRI time series were obtained from 32 transverse slices covering the whole brain with a spatial resolution of 2.5 3 2.5 3 2 mm, with an additional 0.5 mm slice gap, using a single-shot gradient-echo planar sequence (ascending slice order; acquisition matrix 80 × 58; TR = 8000 ms, including 1700 ms acquisition and 6300 ms silent gap; TE = 12 ms; flip angle = 90°).

Quantification and Statistical Analysis

EEG experiment—Two time windows of interest were defined. The first time window between 80 and 120 ms after stimulus onset was chosen to encompass a potential N100 effect reported in human literature^{35–37} and was also examined in Howell and colleagues’¹⁰⁶ auditory oddball paradigm with dogs. The second time window between 300 and 500 ms after stimulus onset aimed to test the N400 effect based on human literature.^{23,38–41,107}

EEG signal for each Subject and electrode was first condition-averaged across trials, then for each time window. Next, the resulting mean amplitudes were entered in the analysis.⁵⁹ We chose to analyze all data using linear mixed models (LMM) – which are more robust to the violations of the ANOVA assumption –, because we also planned an exploratory analysis besides the analysis of the selected time windows based on earlier studies. The exploratory analysis involved multiple models in which we expected that the assumptions of RM ANOVA might be violated. LMMs were implemented by the *lme* function⁹⁹ in R version 1.3.1093.⁹⁸ We performed two separate comparisons:^{7,30} 1) first, we tested the effect of transitional probability by contrasting ERPs for LFs and PWs (equalized for their word frequency), 2) the second comparison tested the effect of word frequency by contrasting ERPs for HFs and LFs (equalized for their transitional probabilities). Predictors in the LMMs included the fixed factors of transitional probability (Model 1) or word frequency (Model 2) and electrode (Fz, Cz), and subject as random factor. Each model included the fixed factors, their interactions and the electrode factor nested in the random factor. To test for the significance of each of the fixed factors, we used the likelihood ratio test as implemented by the *anova* function¹⁰⁰ in R. First, we compared our starting model to a reduced model by removing the interaction term. If models were not significantly different according to the likelihood ratio test, the interaction term was removed from the model for the rest of the analysis. Then, we evaluated the fixed effects by removing each of the fixed factors and testing the reduced model against the model without the interaction term.

An additional exploratory analysis was carried out to identify other possible segmentation-related effects and refine the temporal analysis carried out by the dogs. We repeated the linear mixed effect regression modeling in consecutive 50 ms time windows from 0 (word onset) to 1000 ms that were shifted in steps of 10 ms (e.g., between 0-50 ms, 10-60 ms,

20-70 ms etc., see e.g., Teinonen et al.²³). Following Teinonen and colleagues²³, if an effect was found to be significant (< 0.05 , uncorrected) in at least four consecutive time windows, the consecutive time windows got selected for further analysis. This means that we considered the first significant time window as the onset of the effect and the end of the last significant time window as the offset. We then performed a confirmatory analysis by repeating the mixed effect modeling over these boundaries and then tested the significance of the fixed predictors using the likelihood ratio test as implemented by the *anova* function.¹⁰⁰

Finally, we carried out a follow-up analysis of the word frequency and/or TP effects using a comparison between LFs and NWs, which were built from the same syllables, but differed both in word frequency and TP. We carried out the mixed effect modeling in all previously identified time windows with significant effects, and assumed that the directionality of an effect would be indicative of a word frequency or a TP effect.

For time windows with significant effects we performed visual inspection of residual plots and tested the residuals of the models by Saphiro-Wilk test to ensure they do not reveal any obvious deviations from homoscedasticity or normality (see Figure S1). To rectify an encountered non-normal distribution of the residuals in case of the late TP effect (between 680 and 760 ms) we first used log transformation, then run the *transfom Tukey*¹⁰⁸ and *boxcoxmix*¹⁰⁹ functions of R designed to indicate the correct lambda for data transformation, then we applied the resulting lambda value to the original data. However, none of our data transformation attempts resulted in normally distributed residuals, therefore we confine to a cautious interpretation of this effect. Parameter estimates and their CIs, random effect SD and model fit (Rc and Rm) are provided in Table 2 for each model that resulted in significant effects.

fMRI experiment—Image pre-processing and statistical analysis were performed using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm>). First, functional EPI-BOLD images were realigned using rigid body transformations to correct for head movements. Next, the mean functional image was registered manually to a custom-made individual template anatomical image¹¹⁰ and the resulting transformation matrix was applied to all realigned functional images. The normalized functional data were finally smoothed with a 4 mm (FWHM) Gaussian kernel.

Statistical parametric maps were generated using a linear combination of functions derived by convolving the standard SPM hemodynamic response function with the time series of the conditions. To control for confounds induced by motion in the scanner, a displacement vector was calculated for each frame and entered in the design matrix as an additional regressor of no interest (mean framewise displacement: 0.26, see Power et al.¹¹¹).

To confirm that the two syllable sets were perceived equally by the dogs and they did not affect our principal manipulation (i.e., Structured versus Unstructured), we first performed a whole brain analysis, where we compared the effect of syllable sets heard by the dogs regardless of which condition they were used in. Individual contrast images were computed for each syllable set versus silence, which were then entered in an ANOVA for random

effects group analysis. This analysis yielded no significant cluster either for the comparison of syllable set A > syllable set B or for the opposite direction, confirming that condition differences between the Structured and Unstructured stimuli found in our fMRI experiment were not attributable to systematic acoustic differences between the two syllable sets.

Individual contrast images were then computed for each condition (Structured condition in the baseline and in the test measurement and Unstructured condition in the baseline and in the test measurement) versus silence, which were then entered in an ANOVA for random effects group analysis. All results are reported at an uncorrected voxel threshold of $p < 0.001$, and FWE-corrected at cluster level ($p < 0.05$).

Next, we carried out an ROI analysis based on the Cornell Canine Atlas perisylvian lobe mask.⁶³ To test for both within and between session learning effects, we modeled each audio block from each measurement separately (see Boros et al.⁵⁷ and Gábor et al.⁵⁸ for fMRI adaptation effects in dogs). Trial based parameter estimates (beta values) from each audio block, measurement and condition were extracted from the whole ROI of each dog individually. These parameter estimates were then first averaged within subjects, then between subjects by entering them in a RM ANOVA with condition (Structured and Unstructured), measurement (baseline and test) and repetition (1st, 2nd and 3rd block) as within-subject factors using the *ezANOVA* function¹⁰¹ in R version 1.3.1093.⁹⁸ Assumptions for RM ANOVA were checked using Shapiro-Wilk test of normality and Mauchly's sphericity test (see Tables S2 and S3). Statistical parameters are reported in Table S2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was funded by the Hungarian Academy of Sciences (a grant to the MTA-ELTE "Lendület" Neuroethology of Communication Research Group [LP2017-13/2017]), the Eötvös Loránd University, and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant number 950159). A.D. was supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences, the ú NKP-20-5 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." We thank the Department of Neuroradiology at the Medical Imaging Centre of the Semmelweis University, Budapest and Rita Báji, Barbara Csibra, and Márta Gácsi for training the fMRI dogs. We are grateful to Laura V. Cuaya and Raúl I Hernandez-Pérez for their assistance in the fMRI experiments and Adél Könczöl for her assistance in the EEG experiments. We thank all dog owners and dogs for their participation in the experiments.

Data and code availability

All original data and code has been deposited at figshare and is publicly available as of the date of publication. DOIs are listed in the Key resources table. Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

References

1. Jusczyk PW. How infants adapt speech-processing capacities to native-language structure. *Curr Dir Psychol Sci.* 2002; 11: 15–18.
2. Kuhl PK. Early language acquisition: cracking the speech code. *Nat Rev Neurosci.* 2004; 5: 831–843. [PubMed: 15496861]
3. Swingley D. Contributions of infant word learning to language development. *Philos Trans R Soc Lond B Biol Sci.* 2009; 364: 3617–3632. [PubMed: 19933136]
4. Cole RA, Jakimik J, Cooper WE. Segmenting speech into words. *J Acoust Soc Am.* 1980; 67: 1323–1332. [PubMed: 7372919]
5. White, L. *Oxford Handbook of Psycholinguistics.* Rueschemeyer, S-A, Gaskell, MG, editors. Oxford University; 2018. 5–30.
6. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science.* 1996; 274: 1926–1928. [PubMed: 8943209]
7. Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci.* 1998; 9: 321–324.
8. Saffran JR. Statistical language learning: mechanisms and constraints. *Curr Dir Psychol Sci.* 2003; 12: 110–114.
9. Avarguès-Weber A, Finke V, Nagy M, Szabó T, d’Amaro D, Dyer AG, Fiser J. Different mechanisms underlie implicit visual statistical learning in honey bees and humans. *Proc. Natl Acad Sci USA.* 2020; 117: 25923–25934.
10. Santolin C, Rosa-Salva O, Vallortigara G, Regolin L. Unsupervised statistical learning in newly hatched chicks. *Curr Biol.* 2016; 26: R1218–R1220. [PubMed: 27923125]
11. Takahasi M, Yamada H, Okanoya K. Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata* var. *domestica*). *Ethology.* 2010; 116: 481–489.
12. Chen J, Ten Cate C. Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behav Processes.* 2015; 117: 29–34. [PubMed: 25217867]
13. Kaposvari P, Kumar S, Vogels R. Statistical learning signals in macaque inferior temporal cortex. *Cereb Cortex.* 2018; 28: 250–266. [PubMed: 27909007]
14. Kikuchi Y, Attaheri A, Wilson B, Rhone AE, Nourski KV, Gander PE, Kovach CK, Kawasaki H, Griffiths TD, Howard MA 3rd, Petkov CI. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol.* 2017; 15 e2000219 [PubMed: 28441393]
15. Ramos D, Mills DS. Limitations in the learning of verbal content by dogs during the training of OBJECT and ACTION commands. *J Vet Behav.* 2019; 31: 92–99.
16. Fugazza C, Dror S, Sommesse A, Temesi A, Miklósi Á. Word learning dogs (*Canis familiaris*) provide an animal model for studying exceptional performance. *Sci Rep.* 2021; 11 14070 [PubMed: 34234259]
17. Savage-Rumbaugh ES, Murphy J, Sevcik RA, Brakke KE, Williams SL, Rumbaugh DM. Language comprehension in ape and child. *Monogr Soc Res Child Dev.* 1993; 58: 1–222.
18. Lyn, H. *The Oxford Handbook of Comparative Evolutionary Psychology.* Vonk, J, Shackelford, TK, editors. Oxford University; 2012. 356–378.
19. Gillespie-Lynch, K, Savage-Rumbaugh, S, Lyn, H. *Encyclopedia of Language Development.* Brooks, PJ, Kempe, V, editors. SAGE Publications; 2014.
20. Santolin C, Saffran JR. Constraints on statistical learning across species. *Trends Cogn Sci.* 2018; 22: 52–63. [PubMed: 29150414]
21. Saffran JR, Newport EL, Aslin RN. Word segmentation: the role of distributional cues. *J Mem Lang.* 1996; 35: 606–621.
22. Saffran JR, Wilson DP. From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy.* 2003; 4: 273–284.
23. Teinonen T, Fellman V, Nääätänen R, Alku P, Huotilainen M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 2009; 10: 21. [PubMed: 19284661]

24. Fló A, Brusini P, Macagno F, Nespor M, Mehler J, Ferry AL. Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Dev Sci*. 2019; 22 e12802 [PubMed: 30681763]
25. Milne AE, Petkov CI, Wilson B. Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience*. 2018; 389: 104–117. [PubMed: 28687306]
26. Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. *Infancy*. 2005; 7: 53–71. [PubMed: 33430544]
27. Lu K, Vicario DS. Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc Natl Acad Sci USA*. 2014; 111: 14553–14558. [PubMed: 25246563]
28. Lu K, Vicario DS. Familiar but unexpected: effects of sound context statistics on auditory responses in the songbird forebrain. *J Neurosci*. 2017; 37: 12006–12017. [PubMed: 29118103]
29. Hauser MD, Newport EL, Aslin RN. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*. 2001; 78: B53–B64. [PubMed: 11124355]
30. Toro JM, Trobalón JB. Statistical computations over a speech stream in a rodent. *Percept Psychophys*. 2005; 67: 867–875. [PubMed: 16334058]
31. Wilson B, Slater H, Kikuchi Y, Milne AE, Marslen-Wilson WD, Smith K, Petkov CI. Auditory artificial grammar learning in macaque and marmoset monkeys. *J Neurosci*. 2013; 33: 18825–18835. [PubMed: 24285889]
32. Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Thiele A, Griffiths TD, Marslen-Wilson WD, Petkov CI. Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun*. 2015; 6 8901 [PubMed: 26573340]
33. Karuza EA, Emberson LL, Aslin RN. Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiol Learn Mem*. 2014; 109: 193–206. [PubMed: 24076012]
34. Wilson B, Smith K, Petkov CI. Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *Eur J Neurosci*. 2015; 41: 568–578. [PubMed: 25728176]
35. Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat Neurosci*. 2002; 5: 700–703. [PubMed: 12068301]
36. Ablá D, Okanoya K. Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia*. 2008; 46: 2787–2795. [PubMed: 18579166]
37. Ablá D, Katahira K, Okanoya K. On-line assessment of statistical learning by event-related potentials. *J Cogn Neurosci*. 2008; 20: 952–964. [PubMed: 18211232]
38. Kooijman V, Hagoort P, Cutler A. Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Brain Res Cogn Brain Res*. 2005; 24: 109–116. [PubMed: 15922163]
39. Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res*. 2006; 1123: 168–178. [PubMed: 17064672]
40. Cunillera T, Càmará E, Toro JM, Marco-Pallares J, Sebastián-Gallés N, Ortiz H, Pujol J, Rodríguez-Fornells A. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*. 2009; 48: 541–553. [PubMed: 19580874]
41. Snijders TM, Kooijman V, Cutler A, Hagoort P. Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Res*. 2007; 1178: 106–113. [PubMed: 17931604]
42. McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci*. 2006; 26: 7629–7639. [PubMed: 16855090]
43. McNealy K, Mazziotta JC, Dapretto M. The neural basis of speech parsing in children and adults. *Dev Sci*. 2010; 13: 385–406. [PubMed: 20136936]
44. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci*. 2009; 21: 1934–1945. [PubMed: 18823241]
45. Tobia MJ, Iacovella V, Davis B, Hasson U. Neural systems mediating recognition of changes in statistical regularities. *Neuroimage*. 2012; 63: 1730–1742. [PubMed: 22906790]

46. Durrant SJ, Cairney SA, Lewis PA. Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb Cortex*. 2013; 23: 2467–2478. [PubMed: 22879350]
47. Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang*. 2013; 127: 46–54. [PubMed: 23312790]
48. Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB. The necessity of the medial temporal lobe for statistical learning. *J Cogn Neurosci*. 2014; 26: 1736–1747. [PubMed: 24456393]
49. Graybiel AM. Building action repertoires: memory and learning functions of the basal ganglia. *Curr Opin Neurobiol*. 1995; 5: 733–741. [PubMed: 8805417]
50. Jin DZ, Fujii N, Graybiel AM. Neural representation of time in cortico-basal ganglia circuits. *Proc Natl Acad Sci USA*. 2009; 106: 19156–19161. [PubMed: 19850874]
51. Jin X, Costa RM. Shaping action sequences in basal ganglia circuits. *Curr Opin Neurobiol*. 2015; 33: 188–196. [PubMed: 26189204]
52. Gheysen F, Van Opstal F, Roggeman C, Van Waelvelde H, Fias W. The neural basis of implicit perceptual sequence learning. *Front Hum Neurosci*. 2011; 5: 137. [PubMed: 22087090]
53. Tremblay P, Baroni M, Hasson U. Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage*. 2013; 66: 318–332. [PubMed: 23116815]
54. Farthouat J, Franco A, Mary A, Delpouve J, Wens V, Op de Beeck M, DeTiège X, Peigneux P. Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. *Brain Topogr*. 2017; 30: 220–232. [PubMed: 27613530]
55. Andics A, Gábor A, Gácsi M, Faragó T, Szaboó D, Miklósi Á. Neural mechanisms for lexical processing in dogs. *Science*. 2016; 353: 1030–1032. [PubMed: 27576923]
56. Prichard A, Cook PF, Spivak M, Chhibber R, Berns GS. Awake fMRI reveals brain regions for novel word detection in dogs. *Front Neurosci*. 2018; 12: 737. [PubMed: 30374286]
57. Boros M, Gábor A, Szaboó D, Bozsik A, Gácsi M, Szalay F, Faragó T, Andics A. Repetition enhancement to voice identities in the dog brain. *Sci Rep*. 2020; 10 3989 [PubMed: 32132562]
58. Gábor A, Gácsi M, Szaboó D, Miklósi Á, Kubinyi E, Andics A. Multilevel fMRI adaptation for spoken word processing in the awake dog brain. *Sci Rep*. 2020; 10 11968 [PubMed: 32747731]
59. Magyari L, Huszár Z, Turzó A, Andics A. Event-related potentials reveal limited readiness to access phonetic details during word processing in dogs. *R Soc Open Sci*. 2020; 7 200851 [PubMed: 33489257]
60. Miklósi A, Topál J. What does it take to become ‘best friends’? Evolutionary changes in canine social competence. *Trends Cogn Sci*. 2013; 17: 287–294. [PubMed: 23643552]
61. Pongrácz P. Modeling evolutionary changes in information transfer. *Eur Psychol*. 2017; 22: 219–232.
62. Kis A, Szakadát S, Kovács E, Gácsi M, Simor P, Gombos F, Topál J, Miklósi A, Boódzis R. Development of a non-invasive polysomnography technique for dogs (*Canis familiaris*). *Physiol Behav*. 2014; 130: 149–156. [PubMed: 24726397]
63. Johnson PJ, Luh W-M, Rivard BC, Graham KL, White A, FitzMaurice M, Loftus JP, Barry EF. Stereotactic cortical atlas of the domestic canine brain. *Sci Rep*. 2020; 10 4781 [PubMed: 32179861]
64. Bognár Z, Szabó D, Deés A, Kubinyi E. Shorter headed dogs, visually cooperative breeds, younger and playful dogs form eye contact faster with an unfamiliar human. *Sci Rep*. 2021; 11 9293 [PubMed: 33927315]
65. Bunford N, Hernández-Pérez R, Farkas EB, Cuaya LV, Szabó D, Szaboó Á;G, Gácsi M, Miklósi Á, Andics A. Comparative brain imaging reveals analogous and divergent patterns of species and face sensitivity in humans and dogs. *J Neurosci*. 2020; 40: 8396–8408. [PubMed: 33020215]
66. Dufour S, Brunellière A, Frauenfelder UH. Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cogn Sci*. 2013; 37: 489–507. [PubMed: 23163763]

67. Winsler K, Midgley KJ, Grainger J, Holcomb PJ. An electrophysiological megastudy of spoken word recognition. *Lang Cogn Neurosci*. 2018; 33: 1063–1082. [PubMed: 33912620]
68. Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*. 2015; 88: 2–19. [PubMed: 26447569]
69. Lehéricy S, Benali H, Van de Moortele P-F, Péligrini-Issac M, Waechter T, Ugurbil K, Doyon J. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proc Natl Acad Sci USA*. 2005; 102: 12566–12571. [PubMed: 16107540]
70. Kotz, SA. Max Planck Inst Hum Cogn Brain Sci Leipzig, Habilit thesis. 2006. The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging.
71. Forkstam C, Hagoort P, Fernandez G, Ingvar M, Petersson KM. Neural correlates of artificial syntactic structure classification. *Neuroimage*. 2006; 32: 956–967. [PubMed: 16757182]
72. Friederici AD, von Cramon DY, Kotz SA. Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain*. 1999; 122: 1033–1047. [PubMed: 10356057]
73. Frisch S, Kotz SA, von Cramon DY, Friederici AD. Why the P600 is not just a P300: the role of the basal ganglia. *Clin Neurophysiol*. 2003; 114: 336–340. [PubMed: 12559242]
74. Kotz SA, Friederici AD. Electrophysiology of normal and pathological language processing. *J Neurolinguist*. 2003; 16: 43–58.
75. Kotz SA, Frisch S, von Cramon DY, Friederici AD. Syntactic language processing: ERP lesion data on the role of the basal ganglia. *J Int Neuropsychol Soc*. 2003; 9: 1053–1060. [PubMed: 14738286]
76. Fiser J, Lengyel G. A common probabilistic framework for perceptual and statistical learning. *Curr Opin Neurobiol*. 2019; 58: 218–228. [PubMed: 31669722]
77. Segaert K, Weber K, de Lange FP, Petersson KM, Hagoort P. The suppression of repetition enhancement: a review of fMRI studies. *Neuropsychologia*. 2013; 51: 59–66. [PubMed: 23159344]
78. Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci*. 2006; 10: 233–238. [PubMed: 16616590]
79. Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J Exp Psychol Learn Mem Cogn*. 2005; 31: 24–39. [PubMed: 15641902]
80. Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci*. 2015; 19: 117–125. [PubMed: 25631249]
81. Salthouse TA. Influence of experience on age differences in cognitive functioning. *Hum Factors*. 1990; 32: 551–569. [PubMed: 2074109]
82. Bognár Z, Iotchev IB, Kubinyi E. Sex, skull length, breed, and age predict how dogs look at faces of humans and conspecifics. *Anim Cogn*. 2018; 21: 447–456. [PubMed: 29667089]
83. Gácsi M, McGreevy P, Kara E, Miklósi A. Effects of selection for cooperation and attention in dogs. *Behav Brain Funct*. 2009; 5: 31. [PubMed: 19630939]
84. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper PJ, Lingaas F, Fredholm M, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci USA*. 2012; 109: 8878–8883. [PubMed: 22615366]
85. Virányi Z, Gacsi M, Kubinyi E, Topál J, Belényi B, Ujfalussy D, Miklósi A. Comprehension of human pointing gestures in young human-reared wolves (*Canis lupus*) and dogs (*Canis familiaris*). *Anim Cogn*. 2008; 11: 373–387. [PubMed: 18183437]
86. Topál J, Gácsi M, Miklósi Á, Virányi Z, Kubinyi E, Csányi V. Attachment to humans: a comparative study on hand-reared wolves and differently socialized dog puppies. *Anim Behav*. 2005; 70: 1367–1375.
87. Lenkei R, Újváry D, Bakos V, Faragó T. Adult, intensively socialized wolves show features of attachment behaviour to their handler. *Sci Rep*. 2020; 10 17296 [PubMed: 33057050]
88. Gácsi M, Vas J, Topál J, Miklósi Á. Wolves do not join the dance: sophisticated aggression control by adjusting to human social signals in dogs. *Appl Anim Behav Sci*. 2013; 145: 109–122.

89. Fugazza C, Andics A, Magyari L, Dror S, Zempléni A, Miklósi Á. Rapid learning of object names in dogs. *Sci Rep.* 2021; 11 2222 [PubMed: 33500506]
90. Kaminski J, Call J, Fischer J. Word learning in a domestic dog: evidence for ‘fast mapping.’. *Science.* 2004; 304: 1682–1683. [PubMed: 15192233]
91. Pilley JW, Reid AK. Border collie comprehends object names as verbal referents. *Behav Processes.* 2011; 86: 184–195. [PubMed: 21145379]
92. Fugazza C, Miklósi Á. Depths and limits of spontaneous categorization in a family dog. *Sci Rep.* 2020; 10 3082 [PubMed: 32080273]
93. Griebel U, Oller DK. Vocabulary learning in a Yorkshire terrier: slow mapping of spoken words. *PLoS ONE.* 2012; 7 e30182 [PubMed: 22363421]
94. Kiss, G; Neméth, G; Olaszy, G; Gordos, G. A flexible multilingual TTS development and speech research tool; EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event; September 3-7; Aalborg, Denmark. 2001.
95. Olaszy G, Németh G, Olaszi P, Kiss G, Zainkó Cs, Gordos G. Profivox — a Hungarian text-to-speech system for telecommunications applications. *Int J Speech Technol.* 2000; 3: 201–215.
96. Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C. What’s new in psychtoolbox-3. *Perception.* 2007; 36: 1–16.
97. Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci.* 2011; 2011 156869 [PubMed: 21253357]
98. R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing; 2013.
99. Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Van Willigen B, Ranke J. EISPACK authors; R-core (2021) nlme:linear and nonlinear mixed effects models.
100. Fox, J, Weisberg, S. *An R Companion to Applied Regression.* Third edition. Sage; 2019.
101. Lawrence MA. ez: easy analysis and visualization of factorial experiments. R package version 44-0. 2011.
102. Forgács B, Parise E, Csibra G, Gergely G, Jacquey L, Gervain J. Fourteen-month-old infants track the language comprehension of communicative partners. *Dev Sci.* 2019; 22 e12751 [PubMed: 30184313]
103. Andics A, Gácsi M, Faragó T, Kis A, Miklósi A. Voicesensitive regions in the dog and human brain are revealed by comparative fMRI. *Curr Biol.* 2014; 24: 574–578. [PubMed: 24560578]
104. Kis A, Gergely A, Galambos Á, Abdai J, Gombos F, Bódizs R, Topál J. Sleep macrostructure is modulated by positive and negative social experience in adult pet dogs. *Proc Biol Sci.* 2017; 284 20171883 [PubMed: 29070727]
105. Czeibert K, Baksa G, Grimm A, Nagy SA, Kubinyi E, Petneházy Ö. MRI, CT and high resolution macro-anatomical images with cryosectioning of a Beagle brain: creating the base of a multimodal imaging atlas. *PLoS ONE.* 2019; 14 e0213458 [PubMed: 30845177]
106. Howell TJ, Conduit R, Toukhsati S, Bennett P. Auditory stimulus discrimination recorded in dogs, as indicated by mismatch negativity (MMN). *Behav Processes.* 2012; 89: 8–13. [PubMed: 22001730]
107. Goyet L, de Schonen S, Nazzi T. Words and syllables in fluent speech segmentation by French-learning infants: an ERP study. *Brain Res.* 2010; 1332: 75–89. [PubMed: 20331982]
108. Mangiafico SS. Summary and analysis of extension program evaluation in R, version 1.18.8. 2016. <https://rcompanion.org/documents/RHandbookProgramEvaluation.pdf>
109. Almohaimeed A, Einbeck J. boxcoxm: an R package for response transformations for random effect and variance component models. 2017. <https://cran.r-project.org/web/packages/boxcoxm/vignettes/boxcoxm.pdf>
110. Czeibert K, Andics A, Petneházy Ö, Kubinyi E. A detailed canine brain label map for neuroimaging analysis. *Biol Futur.* 2019; 70: 112–120. [PubMed: 34554420]

111. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*. 2012; 59: 2142–2154. [PubMed: 22019881]

Highlights

- Dogs listened to speech streams with various distributional cues to word boundaries
- ERPs show that dogs track both word frequency and syllable transitional probability
- fMRI reveals syllable sequence processing in dog basal ganglia and auditory cortex
- Dogs use similar neural and computational mechanisms as humans do to segment speech

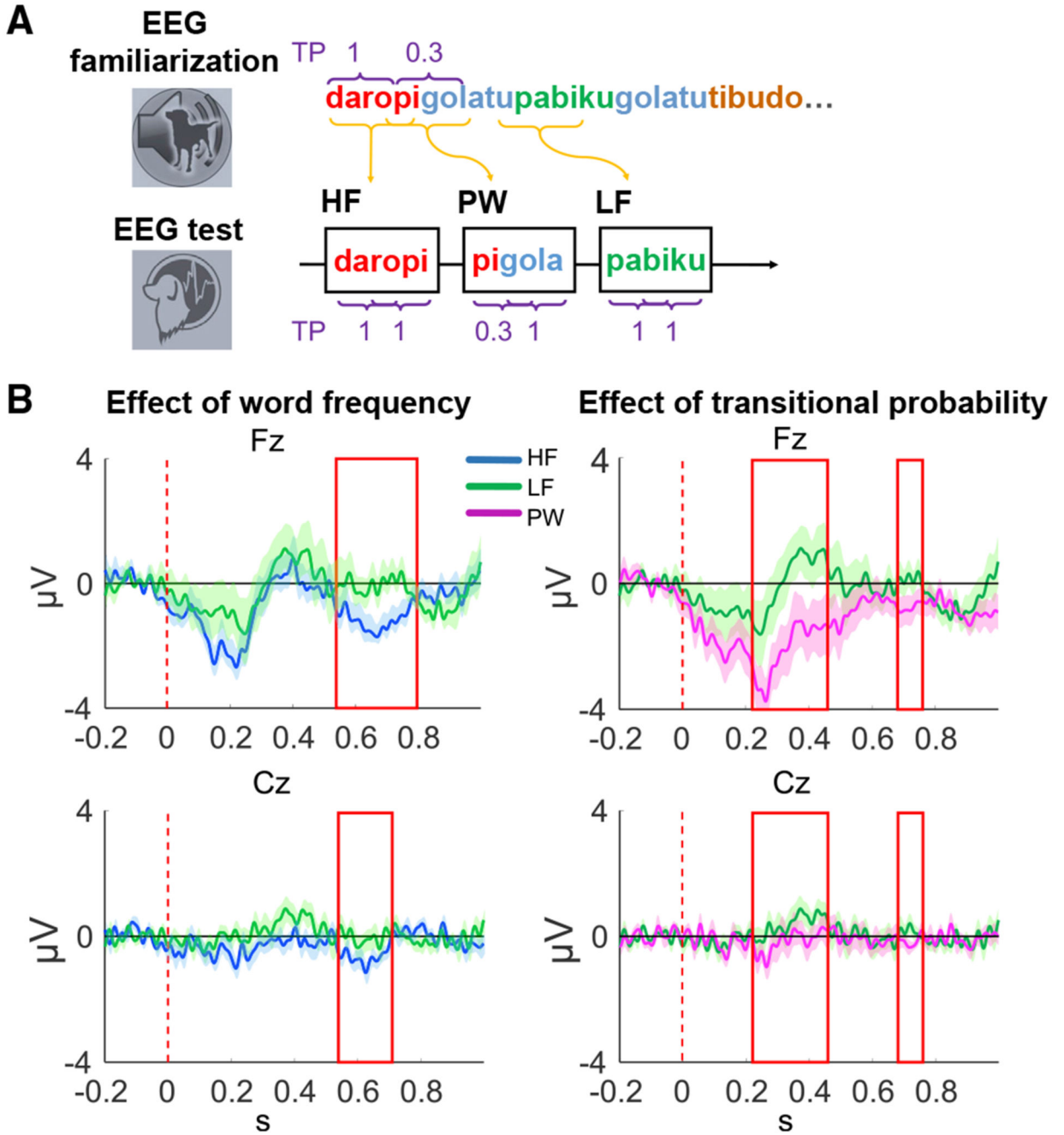


Figure 1. Experimental procedures and results of the EEG experiment

(A) Sessions in the EEG experiment. Top row: familiarization phase is shown. Dogs listened for ca. 8.5 min to a continuous speech stream consisting of a set of 12 syllables forming four artificial words, which followed each other in a random order. Bottom row: test phase is shown. Three conditions were created from the stimuli of the familiarization phase, which were then played back in isolation: high word frequency, high transitional probability words (HF); low word frequency, high transitional probability words (LF); and low word frequency, low transitional probability words (partwords [PW]). TP, transitional probability.

(B) Significant ERP effects of the sliding time window analysis at the Fz and Cz electrodes in the test phase. The effect of word frequency (i.e., frequency of syllable co-occurrence) was obtained by contrasting HF words with LF words. The effect of transitional probability was obtained by contrasting LF words with PWs. Vertical dashed line shows the timing of word onset (0 s), the red rectangle represents the significant time windows where condition differences were found, and shadows represent SEM.

See also Figures S1 and S2 and Tables S1 and S5.

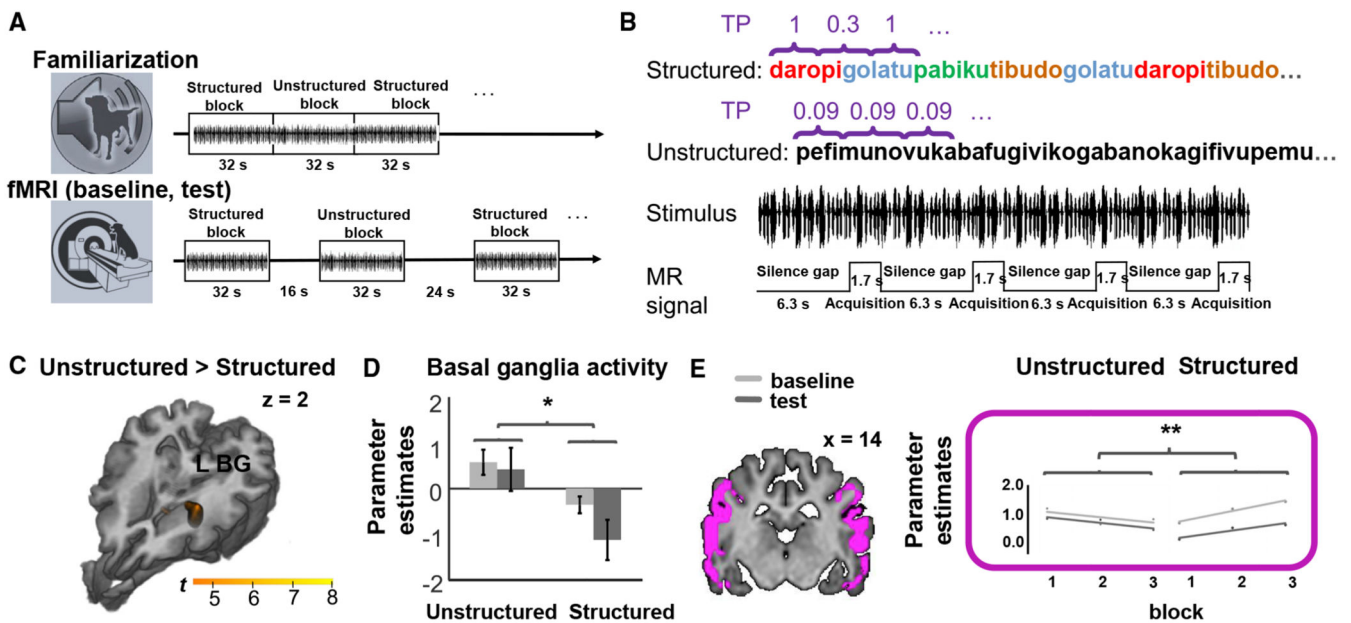


Figure 2. Experimental procedure and results of the fMRI experiment

(A) Sessions of the fMRI experiment: after a baseline fMRI measurement (bottom row), dogs were exposed to the stimuli for 16 min outside the scanner (familiarization, top row) and then immediately underwent a test fMRI measurement (bottom row). The two fMRI sessions were identical and consisted of alternating silence and audio blocks. During the familiarization session, audio blocks followed each other without an inserted silence.

(B) Two types of audio blocks were created. In the structured condition, syllables formed artificial words, characterized by higher transitional probabilities for within-word syllables (TP = 1) than word boundary syllables (TP = 0.33). For the unstructured condition, a different syllable set was used, and the syllables followed each other in a random order, resulting in an overall lower TP for all syllables than 0.3 (mean TP = 0.09). Within one audio block, stimulus presentation was uninterrupted, whereas data were acquired using sparse sampling, i.e., four 1.7-s acquisitions interspersed with silence gaps (6.3 s).

(C) Group level contrasts of the structured and unstructured condition rendered on a template dog brain. The resulting cluster is primarily located at the region of basal nuclei encompassing portions of the caudate nucleus and the thalamus. Thresholds: $p < 0.001$, uncorrected for multiple comparisons at the voxel level and FWE-corrected threshold of $p < 0.05$ at the cluster level. L BG, left basal ganglia. Scale bar represents voxel-level t value.

(D) Parameter estimates of the structured and unstructured condition in the baseline and test fMRI measurements at the main peak of the group level differences $[-6, -8, 6]$. Error bars represent the SEM. $*p < 0.05$.

(E) ROI analysis. Bilateral auditory cortex mask created from the Cornell Canine Atlas⁶³ and the results of the ROI analysis of parameter estimates. $**p < 0.01$. See also Audio S1 and Tables S2, S3, and S5.

Table 1
Characteristics of the conditions in the test phase of the EEG experiment

Condition	Word no.	Words–language 1	Words–language 2	Word frequency	Transitional probability
High-frequency (HF) words	1	daropi	dobuti	high (240)	high (1)
	2	golatu	kubipa	high (240)	high (1)
Low-frequency (LF) words	3	pabiku	piroda	low (120)	high (1)
	4	tibudo	tulago	low (120)	high (1)
Partwords (PWs)	5	pigola	padobu	low (120)	low (0.3)
	6	tudaro	tikubi	low (120)	low (0.3)
Non-words (NWs)	7	dobuti	daropi	0	low (0.11)
	8	kubipa	golatu	0	low (0.11)

Half of the dogs were exposed to speech streams created from words of language 1 and the other half to speech streams created from words of language 2. Transitional probabilities given in brackets represent the probability of syllable 2 following syllable 1 of the trisyllabic artificial word. The transitional probabilities between syllables 2 and 3 did not differ across conditions, except for the NW condition, where it was lower than 0.3 (see also STAR Methods).

Table 2
Summary of the EEG experiment results

Time window	Analysis	Fixed effect	B	CI		Random effect SD		Rm	Rc
				Lower	Upper	Subject	Electrode in subject		
80–120 ms	<i>a priori</i> (N100)	electrode	1.17	0.17	2.17	0.71	0.63	0.08	0.27
300–500 ms	<i>a priori</i> (N400)	TP	-1.31	-2.44	-0.18	0.87	0.00	0.08	0.18
220–470 ms	exploratory	TP	-1.36	-2.45	-0.27	0.94	0.00	0.11	0.23
680–760 ms	exploratory	TP	-0.62	-1.17	-0.07	0.67	0.00	0.06	0.29
540–710 ms	exploratory	WF	0.92	0.31	1.53	0.76	0.00	0.09	0.32
670–790 ms	exploratory	WF 3 electrode	-1.07	-2.04	-0.10	0.57	0.00	0.15	0.34

Table 3
Summary of the whole brain analysis results in the fMRI experiment

Contrast	Region	Z score	Cluster size	p (corr)	Coordinates
Acoustic > silence (overall)	mesencephalon	4.35	30	0.032	8 -18 -2
	L mid ectosylvian gyrus	4.30	26	0.048	-24 -20 20
	L caudal ectosylvian gyrus	3.83			-26 -22 12
	R mid ectosylvian gyrus	3.86	34	0.021	24 -20 10
Unstructured > structured (overall)	L basal ganglia/thalamus	5.08	29	0.016	-10 -6 2
Unstructured > structured (baseline)	no suprathreshold voxels	-	-	-	- - -
Unstructured > structured (test)	L basal ganglia/thalamus	3.91	22	0.048	-6 -8 6

Thresholds: $p < 0.001$, uncorrected for multiple comparisons at the voxel level, and an FWE-corrected threshold of $p < 0.05$ at the cluster level. The table lists all local maxima at least 6 mm apart, for each suprathreshold cluster. No suprathreshold voxels were found for the structured > unstructured comparison.