



Domestic dogs (*Canis familiaris*) recognise meaningful content in monotonous streams of read speech

Holly Root-Gutteridge^{1,2} · Anna Korzeniowska¹ · Victoria Ratcliffe³ · David Reby^{1,4,5}

Received: 12 March 2024 / Revised: 21 January 2025 / Accepted: 26 March 2025
© The Author(s) 2025

Abstract

Domestic dogs (*Canis familiaris*) can recognize basic phonemic information from human speech and respond to commands. Commands are typically presented in isolation with exaggerated prosody known as dog-directed speech (DDS) register. Here, we investigate whether dogs can spontaneously identify meaningful phonemic content in a stream of putatively irrelevant speech spoken in monotonous prosody, without congruent prosodic cues. To test this ability, dogs were played recordings of their owners reading a meaningless text which included a short meaningful or meaningless phrase, either read with unchanged reading prosody or with an exaggerated DDS prosody. We measured the occurrence and duration of dogs' gaze at their owners. We found that, while dogs were more likely to detect and respond to inclusions that contained meaningful phrases spoken with DDS prosody, they were still able to detect these meaningful inclusions spoken in a neutral reading prosody. Dogs detected and responded to meaningless control phrases in DDS as frequently as to meaningful content in neutral reading prosody, but less often than to meaningful content in DDS. This suggests that, while DDS prosody facilitates the detection of meaningful content in human speech by capturing dogs' attention, dogs are nevertheless capable of spontaneously recognizing meaningful phonemic content within an unexaggerated stream of speech.

Keywords Dog-directed speech · Heterospecific communication · Speech recognition · Word recognition · Human-animal communication

Introduction

After more than 14,000 years of domestication, domestic dogs (*Canis familiaris*) have a close and well-developed relationship with humans (Vilà et al. 1997; Thalmann et al. 2013), making them an excellent model for exploring

human-animal communication. Dogs are highly attentive to human speech (Kaminski et al. 2004; Fukuzawa et al. 2005; Adachi et al. 2007; Pilley and Reid 2011; Pilley 2013; Gibson et al. 2014; Ratcliffe et al. 2014; Ratcliffe and Reby 2014; Root-Gutteridge et al. 2019; Boros et al. 2020) and respond to both segmental phonemic cues (Baru 1975 (problematic - welfare validity concerns); Fukuzawa et al. 2005; Ratcliffe and Reby 2014) and emotional prosodic cues (Scheider et al. 2011; Ratcliffe and Reby 2014). Furthermore, dogs can learn to recognise human voices (Adachi et al. 2007; Root-Gutteridge et al. 2019), commands (Mills 2005), and even referential words (Kaminski et al. 2004; Pilley and Reid 2011; Dror et al. 2021; Fugazza et al. 2021). While some of these abilities may use the same acoustic features as when discriminating between conspecific vocalisations (Yin and McCowan 2004; Maros et al. 2008; Taylor et al. 2009; Péter et al. 2014), the extent to which dogs are capable of extracting segmental information that is unique to human speech remains poorly understood. Here, to fill this gap, we explore dogs' ability to identify speech that is meaningful to them among a stream of read speech that is meaningless to them,

✉ Holly Root-Gutteridge
hollyrg@googlemail.com

¹ School of Psychology, University of Sussex, Falmer, East Sussex, UK

² School of Life and Environmental Sciences, University of Lincoln, Lincoln, Lincolnshire, UK

³ Defence Science and Technology Laboratory, Salisbury, Wiltshire, UK

⁴ ENES Bioacoustics Research Lab / Lyon Neuroscience Research Centre (CRNL), University of Saint-Etienne, CNRS UMR5292, INSERM UMR_S 1028, Saint-Etienne, France

⁵ Institut Universitaire de France, Paris, France

while also testing the effects of speaker sex and reading prosody on their performance.

When addressing dogs, humans often use a speech register called dog-directed speech (DDS) which, like infant-directed speech (IDS), is characterized by increased intonation and pitch range, and is thought to attract the listener's attention (Mitchell and Edmonson 1999; Xu et al. 2013; Gergely et al. 2017; Ben-Aderet et al. 2017; Lesch et al. 2019) and make word recognition easier (Thiessen et al. 2005). There is some evidence that adult dogs prefer to associate with people who produce DDS compared to adult-directed speech (Braem and Mills 2010; Jeannin et al. 2017; Ben-Aderet et al. 2017; Andics and Miklósi 2018; Benjamin and Slocombe 2018). Male and female owners are known to exhibit differences in their use and presentation of DDS, with male owners producing less exaggerated speech compared to female owners (Prato-Previde et al. 2006). Furthermore, while dogs respond differently to male and female voices (Ratcliffe et al. 2014; Gergely et al. 2017), the effect of speaker sex on word recognition has not been investigated yet and is thus included in our experimental design. (Prichard et al. 2018; Gábor et al. 2020) Dogs may follow commands more efficiently when presented in DDS (Mills et al. 2005). Yet the specific contribution of speech prosody to dogs' perception of speech signals is not well understood, and, in particular, it remains to be established whether DDS prosody is essential for command recognition. To test this, we included prosody type as a condition in our experimental design.

Speech is typically presented as a long stream of phonemes, often against background noise provided by either environmental sound or other speakers. A subset of human speech research has examined the "cocktail party effect", first identified in humans by Cherry (1953), where speech of interest is extracted from background "babble", and attended to while the "babble" is discarded (Gábor et al. 2020). Humans can separate these streams and pick out the salient speech stream against multi-talker babble (Cherry 1953). There is evidence that dogs are capable of recognising their name when presented in a multi-speaker babble, even outperforming one-year-old human infants, provided the loudness of meaningless-to-them speech and their own name were of equal intensity (Mills et al. 2005). The dogs were shown to attend to their names by orienting towards the sound source (Mallikarjun et al. 2019). This suggests that they have some ability to recognise familiar content when set in noise.

Similar to the ability to separate speech from noise, language comprehension also requires the ability to parse streams of speech into meaningful phonemic units, a process known as speech segmentation (Thiessen et al. 2005). Because speech segmentation is a key part of language

acquisition (Thiessen et al. 2005), it has received much scrutiny in humans (Jusczyk et al. 1999; Mattys et al. 1999; Thiessen et al. 2005; Bortfeld et al. 2005; Seidl and Johnson 2006; Bergelson and Swingley 2012; Vihman 2017; Westermann and Mani 2017). Human infants can perform this speech or word segmentation from birth (Mandel et al. 1995; Fló et al. 2019), and by 6 months, they can recognise target words if their name precedes them (Bortfeld et al. 2005). Infant-directed speech is used to emphasise word boundaries through exaggerated enunciation and increase attentiveness through increased modulation of tone, which may aid the comprehension of speech and emphasise word boundaries, characteristics also seen in dog-directed speech (Burnham and Francis 1998). The ability to segment speech partially depends on statistical learning of the probability that certain sequences of phonemes are more likely than others, allowing the prediction of word boundaries. For example, Fló et al. (2019) claim that *bana...* is likely to be followed by "na", but "*banana*" can be followed by "split, peel, republic", and thus the parsing of the sequence beyond the trisyllabic word is more difficult. Therefore, listeners are more likely to succeed at parsing complex utterances if the speech presented is familiar and the phonemes follow known statistical distributions.

Beyond statistics, exaggerated prosody, which can emphasise pauses and breaks between words, can give a further cue to speech segmentation. Floccia et al. (2016) found that young infants' ability to segment speech depended on the use of exaggerated, infant-directed speech prosody, and discussed the impact of prior experience of speech prosody on word recognition. However, Fló et al. (2019) determined that even newborn humans do not require prosodic cues to recognise words. Previous research has shown that speech with flat fundamental frequency is perceived by human listeners to be less intelligible than more modulated speech (Brokx and Nooteboom 1982; Laures and Weismer 1999; Binns and Culling 2007). While some studies have found that prosodic cues are less useful to adult listeners learning new languages than to infants (Gerken 1996; Matzinger et al. 2021), this finding was not upheld in other studies where prosody was found to be important to adult speech segmentation and learning (Valian and Levitt 1996; Kim et al. 2012; Spring et al. 2013). Overall, learning to segment speech is supported by both exaggerating speech prosody and the pauses between words, but these pauses are rare in adult-directed speech (Matzinger et al. 2021). Other factors influencing speech perception include the amplitude of speech, the speaker's familiarity, and their sex (Cherry 1953; Fant and Tatham 1975; Kuhl 1988; Childers and Wu 1991; Nygaard and Pisoni 1998; Binns and Culling 2007; Johnson 2008). Thus, it can be predicted that familiar speech presented with highly modulated frequency and increased

prosody will gain more attention and be easier to segment from streams than more monotonous speech, but that segmentation of monotonous speech should still occur. Here, we examine whether brief speech pauses are sufficient for dogs to parse familiar, meaningful content from unfamiliar speech streams.

Dogs constitute an excellent model for comparative research into speech perception and heterospecific listening due to their near-constant exposure to human speech and their own interest in following our cues (Taylor et al. 2014). Dogs are commonly exposed to complex speech and dog owners embed meaningful phrases in longer, non-salient speech streams when addressing their pets, suggesting that they expect that their dogs are capable of speech-recognition (Mitchell and Edmonson 1999). Indeed, dogs appear to possess the neural architecture to support it (Boros et al. 2021) and exhibit different processing responses when hearing either familiar or unfamiliar words, supporting the idea that they learn words, rather than just intonational cues or physical gestures (Prichard et al. 2018; Gábor et al. 2020). However, it has been shown that dogs' ability to recognise words falls when ambiguity increases, such as when commands are embedded in more complex sentences or following pauses (Braem and Mills 2010), and it is not clear how far the speech register affects their ability to detect (or their propensity to respond) to meaningful words. Researchers used a combination of fMRI to investigate dog brain structure and EEG scans to investigate whether dogs have the neural mechanisms required to parse speech (Boros et al. 2021). They familiarised the dogs with a stream of artificial speech, with different frequencies of word repetition and paired words and then tested the dogs' event-related responses (ERPs) to the presentation of the learnt speech (Boros et al. 2021). They found that the dogs showed stronger responses to the high frequency artificial words than the low-frequency artificial words, suggesting that dogs segmented the speech into the known syllables and were capable of statistical learning of the co-occurrence of speech syllables (Boros et al. 2021). However, it remains to be established whether dogs express this ability when processing normal speech and in their normal, everyday behaviour.

Dogs produce sequences of barks which can be separated by pauses or continuous, "fused" barks which overlap (Yin and McCowan 2004). However, understanding which units of these sequences are meaningful and decoding them is challenging as it requires pairing call with context (Yin and McCowan 2004), which can be challenging to determine. However, there is evidence from dingoes (*Canis familiaris dingo*) that multi-segment syllables are combined to form novel meanings (Deaúx et al. 2016). However, it is difficult to establish how far dogs parse the sequences of barks or respond to familiar content within sequences.

Using human speech overcomes this issue as it is already established which the meaningful units are and what the response should be. Furthermore, it has been suggested that domestication has shaped the speech perception abilities of species (Hare et al. 2002), thus they may be more strongly pronounced in the species which are accustomed to hearing human speech, including dogs. We suggest that dogs are a good model for exploring how well animals recognise call units within sequences as they produce call sequences and have had long-term exposure to human speech as meaningful. Here, we explore whether dogs are able to discern the presentation of meaningful content in a stream of meaningless-to-them speech and therefore their ability to parse sequences to identify salient phrases.

In order to investigate whether dogs can detect meaningful speech phrases embedded in a stream of speech, and how important tone cues are to their speech recognition we tested dogs' ability to recognise a meaningful phrase ("[dog's name], come on then") within a longer speech stream, presented in either DDS speech or neutral-tone speech. We assessed whether dogs exhibited attention to their owner on hearing the salient content in (a) DDS and (b) neutral reading prosody (NRP) voice speech. We predicted that while dogs may be able to respond to speech either containing meaningful content or presenting DDS prosody, responses should be strongest to speech combining meaningful content pronounced using dog-directed speech. Finally, we investigated whether dogs would either respond more readily to NRP presented by male owners as compared to female owners or show less differentiated responses to men's DDS and NRP speech.

Materials and methods

To test whether dogs would respond more to meaningful than non-meaningful content pronounced in the same speech register, we first piloted their response to speech in Dog-Directed Speech register (DDS, Pilot Experiment), then tested their response to their owners' speech in Neutral Reading Prosody (NRP) vs. DDS (Experiment 1: Prosody), and explored whether owner sex modulated the response of dogs in these tasks (Experiment 2: Sex).

Stimuli

70 owners were recorded reading aloud one of three short (15–20 s) passages from the standard oral reading passage "the rainbow passage" (Fairbanks 1960), with the test phrases produced after 7–12 s as part of the text. The non-meaningful (control) phrases were "[Alfie / Bertie], pass me a coffee!" and the meaningful phrase was "[Dog's

name], come on then!”, chosen as these words had the highest frequency of use by English-speaking owners during interactions with their dogs and were therefore likely to be meaningful to all dogs (Mitchell and Edmonson 1999). The duration of the target phrases was between 0.7s and 2.5s (mean = 1.4s, std. dev. = 0.2), depending on the speaker’s natural talking speed and the number of syllables in the dog’s name. In total, three different extracts of the same length were used and the phrases were included within the sentences, i.e., “There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say, [*Bertie, pass me a coffee*] / [*Dog’s name, come on then*], he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways.” (See ESM for extracts 2 and 3.) The time it took the owners to reach the included phrase depended on the speed of their natural speech (mean = 8.7s, std. dev. = 1.2s) but was consistent across readings by the same individual.

The choice of extract was randomised but if the dog had a name too similar to Alfie or Bertie, they were given an abstract that presented the non-similar name (e.g., the participant dogs Betty and Beans heard the extract which contained Alfie, not Bertie as the control name). For each dog, the same extract was used for all conditions. Voice recordings were made on a Zoom H4N-Pro handheld recorder (Zoom) in a sound-proof booth on campus at University of Sussex. To avoid habituating the dogs to the speech, owners were recorded reading the passages without the dog present in the recording booth and were asked to imagine they were speaking to the dog. Owners were asked to produce the target phrases in (a) their normal reading voice prosody (NRP) and (b) dog-directed speech prosody (DDS). There was an expectation that the DDS speech would show increased pitch and range compared to NRP and that this would be more interesting to the dogs (Lesch et al. 2019). Thus, two recordings were made for the Pilot Experiment: DDS-meaningful and DDS-control; four recordings were created

by each owner for the main experiments: NRP-meaningful, NRP-control, DDS-meaningful, and DDS-control.

All the voice recordings were cut and aligned using the sound software Audacity (Mazzoni and Dannenberg 2015) and the amplitude normalized to -9dB. Mean and coefficient of variation of fundamental frequency (f_0 CV = (f_0 standard deviation / f_0 mean) * 100) were measured in Praat (Boersma and Weenink 2009). f_0 CV provides a standardised measure of f_0 variability independent of f_0 height that takes perception into account (i.e., a modulation of 10 Hz around 100 Hz is perceptually equivalent to a modulation of 100 Hz around 1,000 Hz). Values are presented in Table 1. Within sexes, mean f_0 differed significantly between target phrases (female: $F_{3,108} = 68.3, p < 0.001$; male: $F_{3,43} = 43.0, p < 0.001$). However, mean f_0 did not differ significantly for control vs. meaningful within DDS and NRP registers (LMM: $p > 0.05$ for all), but DDS and NRP phrases did differ significantly from each other ($p < 0.001$ for all). For male owners, coefficient of variation did not differ between any comparison ($p > 0.25$ for all comparisons, $F_{3,43} = 0.6, p = 0.592$ overall). For female owners, coefficient of variation did differ significantly overall ($F_{3,127} = 5.1, p = 0.002$) but only for DDS-Meaningful to all others ($p < 0.010$ for all DDS-meaningful pairwise comparisons), while other pairwise comparisons were non-significant at $p > 0.2$ for all. Thus, DDS speech differed to NRP speech but phrases within speech registers were not significantly different except for female coefficient of variation.

Linear mixed models were applied to a subsample of (a) 5 female and (b) 5 male voices and were used to confirm that the mean fundamental frequency and coefficient of variation of the included phrases presented in NRP did not differ significantly from that of the rest of the read speech. (Mean fundamental frequency LMMs: female - $F_{1,14} = 2.9, p = 0.108$; male - $F_{1,14} = 0.6, p = 0.449$. Coefficient of variation LMMs: female - $F_{1,14} = 2.4, p = 0.141$; male - $F_{1,14} = 2.9, p = 0.111$.)

Participants

Fifty-three privately-owned dogs were recruited through Facebook adverts, flyers, and personal contacts, and tested in a designated testing room on campus at University of Sussex. A total of 57 owners (17 male, 40 female) participated, with a maximum of 3 dogs per owner. (As experiment 2 examined the effect of owner sex, the dogs’ response to both their male and female owners was tested, leading to more owners than dogs.) Trials were discarded if the dog was distracted by non-stimuli sounds or events, e.g., background noise ($n = 1$), the dog was barking continuously ($n = 1$), or if they moved out of camera shot ($n = 2$). We retained data from 49 dogs (24 females and 25 males), from 39 breeds

Table 1 Mean and standard deviation of mean and coefficient of variation of fundamental frequency for (a) the target phrases produced by all speakers and (b) the target phrase and entire speech of 10 speakers

Sampled speech	Speech	Mean f_0 (Hz)	Coefficient of variation
All speakers’ target phrases across all conditions	DDS-meaningful	375 +/-139	14 +/-8
	DDS-control	346 +/-146	16 +/-6
	NRP-meaningful	158 +/-36	12 +/-4
	NRP-control	162 +/-34	13 +/-5
10 speakers’ NRP speech for target phrase vs. read speech	NRP-target phrase	153 +/-38	15 +/-5
	NRP-background speech	159 +/-40	14 +/-4

and cross-breeds, aged between 9 months and 12 years old (mean = 4.1 years, SD = 2.9 years) in our analyses (see ESM Table 1 for details following Volsche et al.'s (2023) suggested format).

Protocol

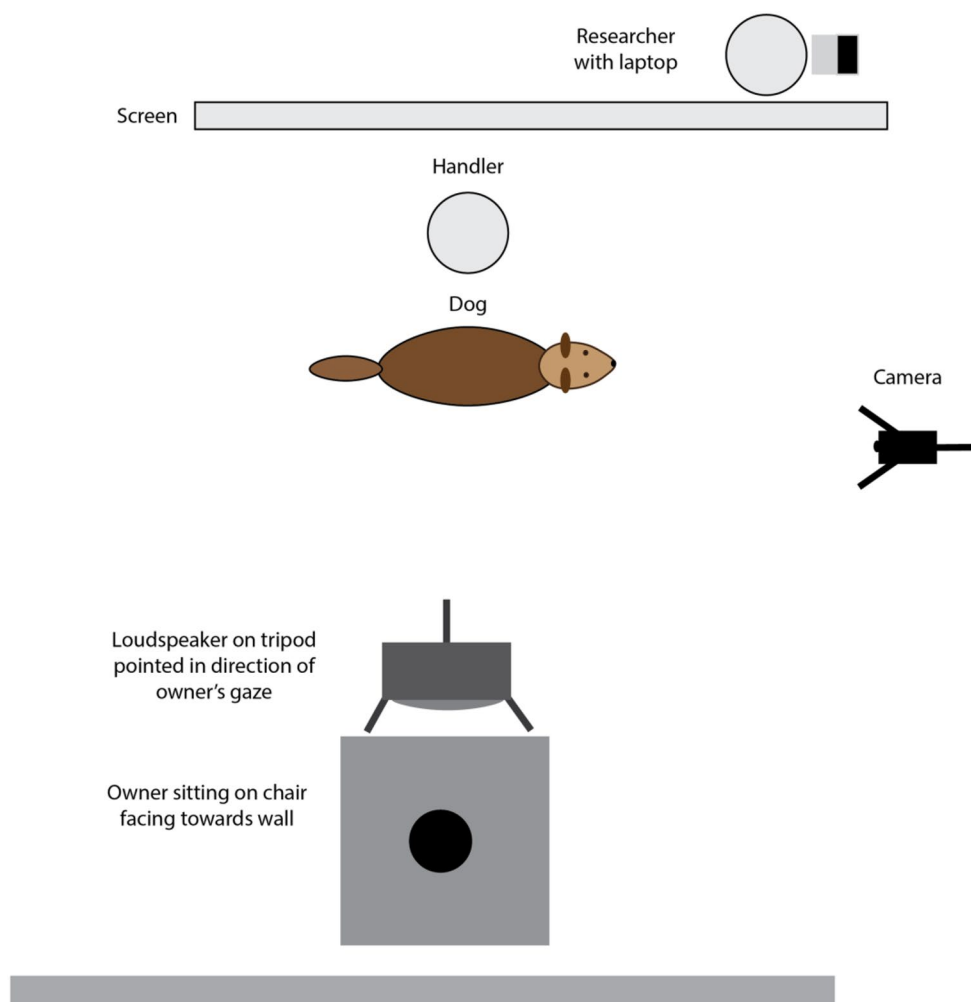
Dogs were introduced to the room and given up to 20 min to freely explore and habituate to the space. Trials began once the dog was considered to be relaxed based on the owner's assessment and the dog's behaviour, e.g., the dog adopted a rest posture such as a sit, they were not panting, barking, whining, or attempting to access the owner. No dogs appeared to be stressed either before, during, or after the trials, using the signs of stress first marked by Beerda et al. (1997), e.g., panting, whining, or circling.

During all trials, the owners wore noise-cancelling headphones (TaoTronics TT BH-047) and listened to music while seated in a chair at 90 degrees to the dog (Fig. 1), with their back to the dog and instructed not to turn to look at the dog. A single Behringer Europort MPA40BT-PRO speaker

was set on a tripod behind the owner's head and set to conversational volume (approx. 65dB measured at the dog's position). The experimenter stood out of the dog's sight line and controlled the stimuli from an Apple MacBook Pro. The dogs were held on a loose lead by the handler and allowed some freedom of movement. While the handler was consistently one of two researchers, their familiarity to the dog could vary from "completely unfamiliar" to "person the dog met on more than one occasion but do not have a close relationship to" if the dog had participated in a previous study before or belonged to a friend of the researchers.

The dogs were positioned either to the left or the right of the speaker, and this position was counter-balanced across dogs within experiments, with half to the left and half to the right. The dogs' reactions were filmed on a Sony FDR-AX100 camcorder (Sony) on a tripod positioned approximately 1.5–2 m from the dogs' starting position. The inter-trial interval depended on the dogs' disposition. If the dog was calm, e.g., resting in one position, such as lying down, and not vocalising or attempting to attract the owner's attention, the trial interval was less than 2 min. However, if

Fig. 1 Experimental set-up in testing room at University of Sussex with the speaker positioned to the dog's left. In half of the trials, this arrangement was reversed with the speaker positioned to the dog's right. The owner was seated facing away from the dog wearing headphones and listening to music while the dog was positioned behind their chair and held on a loose lead by a handler. The speaker was positioned behind the owner's head to simulate them speaking



the dog was restless or distracted, e.g., roaming around the room, vocalising, or focusing intently on sounds or scents in the room, a short break of a few minutes was provided, and the dog was sometimes taken out of the room and returned.

As some owners brought more than one dog and some dogs heard more than one owner, we considered each pairing of owner and dog to be a unique dyad, and thus the unit of comparison was dyad not owner or dog.

Whether the dogs gazed at their owner or not in the 10s period following the included phrase was used as the broadest metric of attention, while duration of gaze was used as the index of attention. None of the dogs were looking at or fully oriented towards their owner immediately prior to the onset of the target phrase, which would have been a criterion for dropping the trial. The trial ended 10s after the onset of the included phrase. Throughout the trials, the handler stood still if the dog was still or followed them if they approached the owner after target phrase (no dog approached before the target phrase).

Pilot experiment: the effect of meaning on dogs' responses to content presented in dog-directed speech (DDS) prosody

The pilot experiment was designed to test whether dogs responded differently to inclusions containing meaningful phrases vs. meaningless, control phrases, in both cases spoken with dog directed prosody (DDS). If they did not respond to the DDS presentation of speech, it was felt that it was unlikely that they would do so to NRP speech and that a new protocol would be required. Twenty-two dogs were tested, and 40 trials from 20 dogs were retained, with 2 dogs removed because they moved out of camera view during the stimulus. All owners included in this experiment were female. Each dog was presented with a recording of their female owner reading the text twice, once including the meaningful phrase and once including the control phrase. The order of presentation of meaningful and control phrase recordings was counter-balanced across dogs.

Experiment 1 prosody: impacts of prosody and content on response

To better explore the effects of prosody and content, the pilot protocol was repeated with a total of 43 owner-dog dyads and all four speech conditions, adding NRP-meaningful and NRP-control to the DDS versions. The dogs heard all four speech conditions in pseudo-randomised presentation, counter-balanced across dogs. A total of 172 trials were retained (13 dogs heard 8 trials, with 4 trials from their male owner and 4 trials from their female owner, but one of these dogs moved out of shot).

Experiment 2 sex: the effects of sex on dogs' responses to content and prosody

During initial data collection, it was noted that some of the dogs appeared to be more responsive to the male owner's NRP speech than their female owner's NRP speech. Therefore, we decided to explore the potential effects of speaker sex on their responses. Thus, we tested whether dogs hearing both their male and female owners would respond differently to them across all four conditions of meaning and prosody, with an expectation that NRP from male owners could elicit more or stronger responses than female NRP due to the smaller differences between male NRP and DDS.

Each of the 13 dogs heard a total of 8 trials, 4 from each owner. To avoid the possible effect of learning on response to the target phrases, as the same text passage was used throughout, the NRP trials were always played first for each owner, with control and meaningful phrase presentation cross-balanced within DDS conditions. Both owners were present in the room, but the non-participant (e.g., the male while the female was "talking" to the dog) was kept out of view to prevent any "clever Hans" effect influencing the results.

One dog was removed from the dataset because he moved out of camera shot while reacting to his owners' voices and thus his responses could not be coded. Another dog (Emma, terrier) had been previously tested in the pilot experiment with a gap of several months between tests, but all other dogs experienced the stimuli as a novel presentation and it was expected that Emma would not retain her memories of the pilot experiment or be primed by them. (Most studies of memory in dogs focus on timespans measured in minutes (Fiset et al. 2003; Fugazza et al. 2016), but there is some evidence they can retain memories for a year or more for trained tasks (Lazarowski et al. 2021), but this does not address non-reinforced experience. However, we note that Emma may represent an outlier in the data.) Thus, 96 trials were retained from 12 dogs in total, with each dog hearing a total of 8 trials, including all four speech presentations from both their male and female owners.

All eight trials were performed on the same day and between trial intervals varied from a few minutes to more than 20 min depending on the behaviour of the dog, e.g., engagement in other activities like sniffing or investigating the area. We counterbalanced the presentation of male and female owners' speech, but each dog heard all four trials from each owner as a block which was not divided (e.g., male owner trials x 4 then female owner trials x 4, but not male owner x 2 then female x 2 etc.). The dogs heard the same order of presentation for both male and female owners (e.g., 1) NRP-meaningful, 2) NRP-control, 3) DDS-meaningful,

then 4) DDS-control for both owners, cross-balanced order across dogs) to avoid order effects on their responsiveness.

Behavioural analysis

Prior to analysis, the videos of the trials were edited in iMovie (Apple Inc.) so that each file presented a single trial with a sound effect replacing the target phrase. All videos were blind coded in Sportscode Gamebreaker 11 (Sportstec, Warriewood, NSW, Australia) by HRG and 25% were second-coded by ATK. Response was defined as the dog directing its gaze towards the owner. The binary gaze response following presentation of the target phrase and duration of response were recorded for each trial. The duration of reaction was capped at 10 s, which was the maximum length of speech measured from the start of the target phrase to the end of the extract. Inter-observer rating agreement was measured for binary gaze and duration using Cronbach's alpha. This resulted in a score of 0.98 out of 1 for binary gaze and 0.94 out of 1 for duration of gaze, which is considered excellent (Bland and Altman 1997).

All coding data files are available as Excel files in the electronic supplementary material and available on Dryad (<https://doi.org/10.5061/dryad.stqjq2c1s>).

Statistics

All statistics were performed using SPSS (SPSS Inc., Chicago, IL., USA) v. 25 and v. 27. Binomial generalized linear mixed effect models (GLMMs) were applied to the binary gaze response variable to examine the effect of tone and

content on gaze response. Duration of response was measured in trials where dogs demonstrated a gaze response, and normal-distribution linear mixed effect models (LMMs) fitted with restricted-maximum likelihood estimation were used to examine the effect of tone and content on gaze duration. For all models, dog identity was included as a random effect and fixed effects were meaning & speech register (DDS or NRP), dog age, and sex. In Experiments 1 and 2, speaker sex was also included as a fixed effect. For experiment 2: Sex, the trial number was included as a random effect. Data were checked for violations of GLMM and LMM assumptions and were not found to be in violation, with residuals normally distributed for the LMM. Bonferroni corrections were used for the p-values throughout, set according to the number of comparisons. Only full models were considered with no stepwise selection undertaken.

Results and discussion

Full results of all models below are presented in Electronic Supplementary Tables 3–8.

Pilot: the effect of meaning on dogs' responses to content presented in dog-directed speech (DDS) prosody

In the pilot experiment, where only DDS phrases were presented, the dogs reacted to the phrase inclusions in 87.5% of trials (35 out of 40), only showing no response in 4 DDS-control trials and 1 DDS-meaningful trial (Table 2; Fig. 2a).

Table 2 Count of trials with positive gaze at owner and the mean and standard deviation of duration of positive gaze at owner response for each experiment by phrase, with only positive responses included for gaze duration

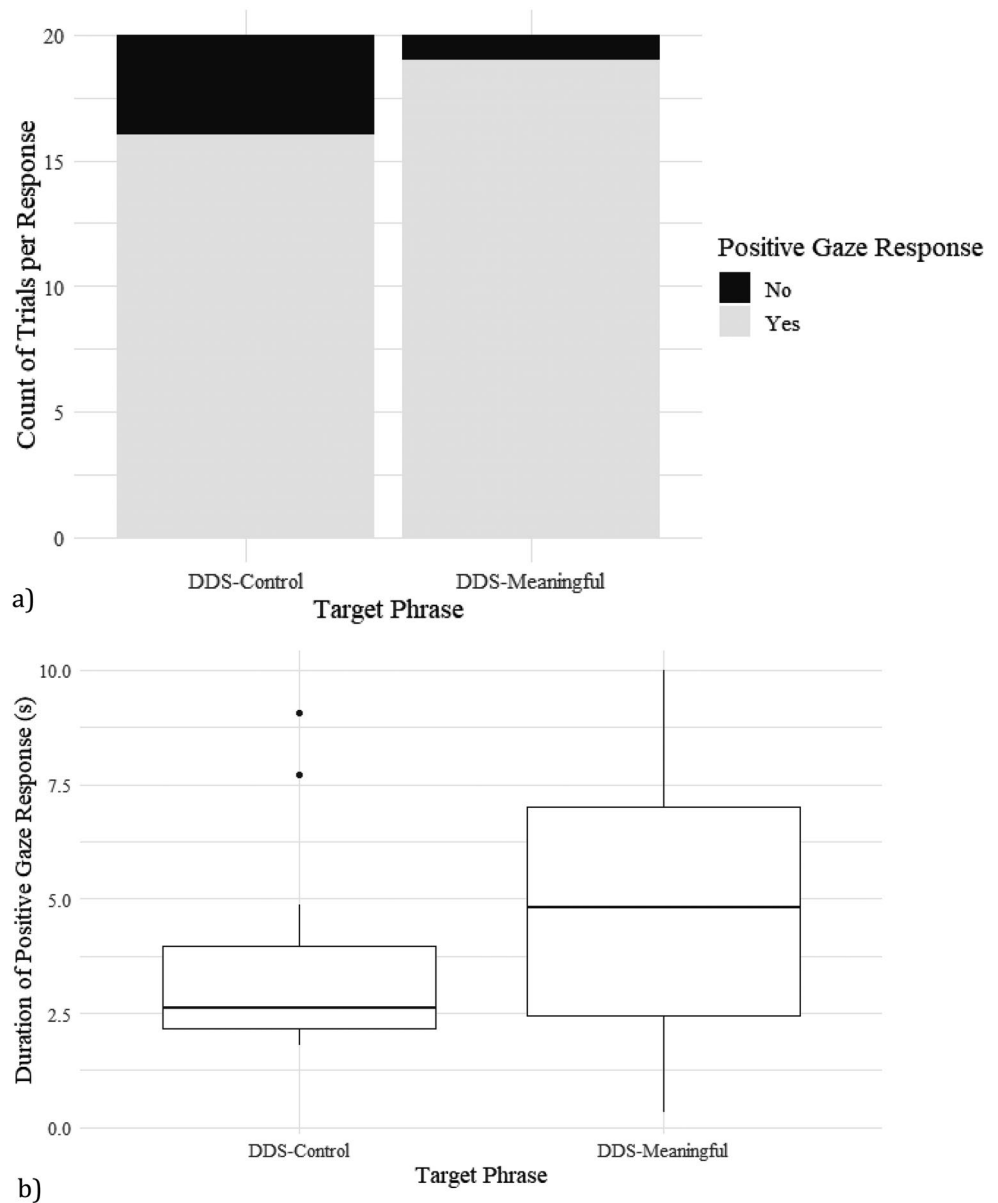
Study	Variable	NRP-control	NRP-meaningful	DDS-control	DDS-meaningful
Pilot	N of positive gaze / total trials	-	-	16 / 20	19 / 20
	Duration (s) Std. Dev.	-	-	3.49 (2.13)	5.11 (2.98)
1	N of positive gaze / total trials	5 / 43	26 / 43	26 / 43	42 / 43
	N of dogs showing positive gaze / total dogs tested	5 / 31	26 / 31	20 / 31	30 / 31
	Duration (s) Std. Dev.	1.72 (1.1)	2.37 (1.1)	3.09 (2.2)	3.92 (2.5)
2	N of positive gaze / total trials	2 / 24	15 / 24	17 / 24	24 / 24
	N of dogs showing positive gaze / total dogs tested	2 / 12	10 / 12*	11 / 12**	12 / 12
	Duration (s) Std. Dev.	1.06 (2.0)	2.46 (1.3)	2.68 (1.6)	3.35 (2.2)

NB: In Experiment 2 Sex, the dogs were given 2 opportunities to respond to the sound of their owner's voice in each condition. Thus, we noted whether they responded differently to male and female owners thus

* 2 dogs did not respond to the NRP-meaningful phrase from either owner, 2 responded only to the female owner and 3 only to the male owner

**1 dog did not respond to the DDS-control phrase from either owner, 2 responded only to the female owner and 3 only to the male owner

Fig. 2 Pilot Experiment: **a)** Stacked bar chart of binary gaze towards owner by dogs in response to target phrases spoken in DDS. Target phrase did not have a significant effect on binary gaze response ($F_{1,29} = 3.0, p = 0.096$). **b)** Boxplot of duration of gaze response to the phrases, where target phrase had no significant effect ($F_{1,16} = 3.3, p = 0.087$)



(NB: The one dog that did not respond in either trial showed no obvious signs of hearing impediment and responded readily to his name outside of the experiment.) The binary gaze response variable was not significantly affected by speech content (binomial GLMM: $F_{1,36} = 2.1, p = 0.155$), dog age ($F_{8,29} = 0.0, p = 1.000$), or dog sex ($F_{1,29} = 0.3, p = 0.601$).

The LMM revealed no significant effect of speech content ($F_{1,16} = 3.3, p = 0.087$), dog age ($F_{8,9} = 1.2, p = 0.412$) or sex ($F_{1,8} = 1.0, p = 0.347$) on duration of positive gaze response once the response was elicited (Fig. 2b). 95% confidence intervals for DDS-control were lower bound=2.7s and upper bound=5.5s, vs. for DDS-meaningful 4.3s and 6.8s respectively, mean difference=1.5s.

It was concluded that dogs responded to DDS without other cues from their owner when played from a speaker but

that this was not influenced by speech content. Therefore, further tests of responsiveness to speech were undertaken.

Experiment 1 prosody: effects of prosody and meaningful content on response

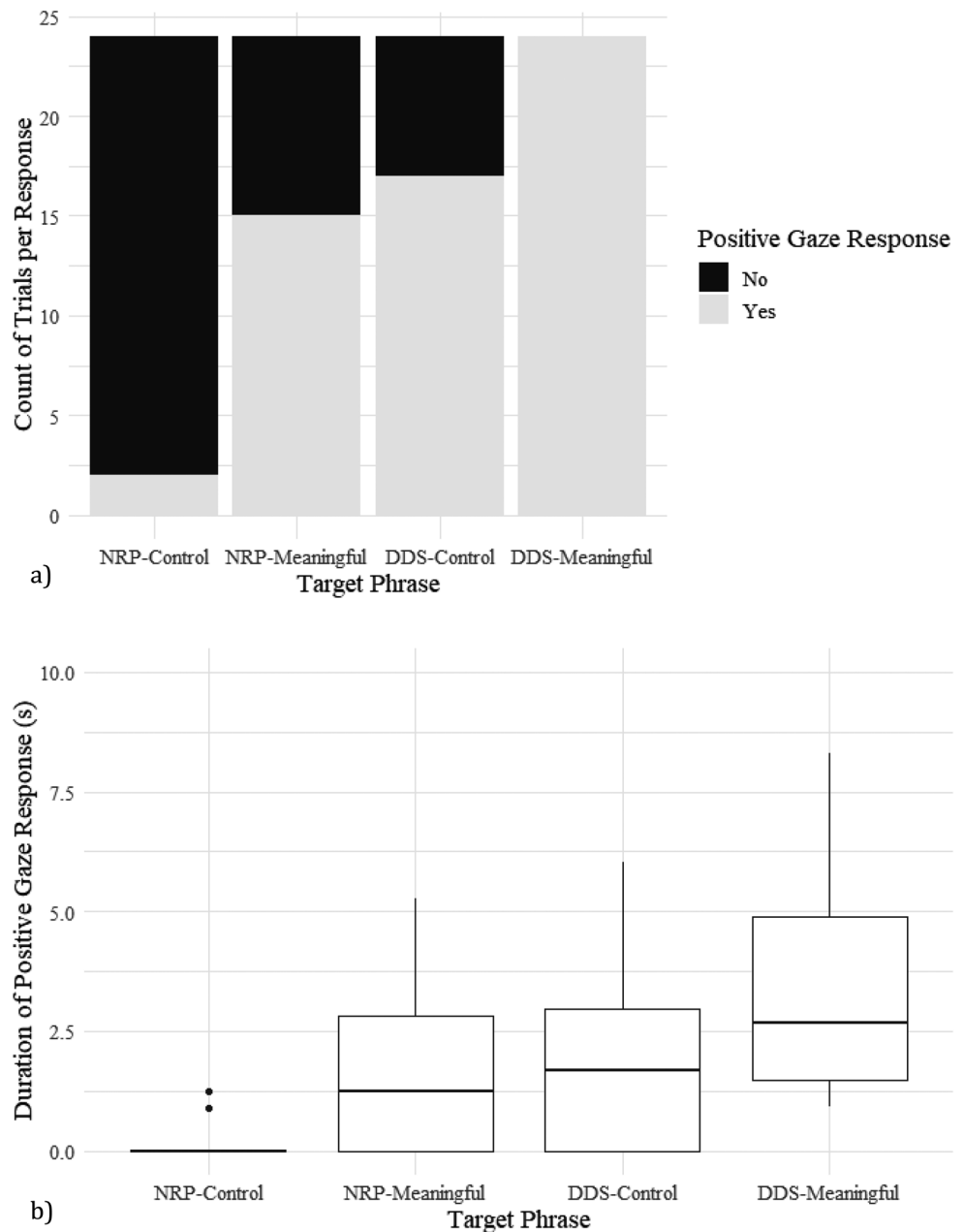
For Experiment 1: Prosody, where we examined the dogs' overall responses to prosody and meaningful content, the dogs showed a positive gaze response to their owners in 57.6% of trials (99 of 172, Table 2). Due to multiple comparisons, we used a Bonferroni correction and significance was $p < 0.017$. The GLMM showed that there was a significant effect of speech on the binary gaze response ($F_{3,165} = 12.3, p < 0.001$), but there was no significant effect of owner sex ($F_{1,165} = 4.0, p = 0.046$), dog age ($F_{1,165} = 0.7, p = 0.390$),

or dog sex ($F_{1,165} = 1.2$, $p=0.272$). Post-hoc tests showed that all pairwise comparisons of reactions to the four phrase-types were significant at $p<0.001$ except between NRP-meaningful and DDS-control ($p=1.000$, see ESM Table 5 for full pairwise comparisons). Overall, dogs were most likely to look at their owners in the DDS-meaningful condition, and least likely in the NRP-control condition (Fig. 3a).

The LMM showed that duration of gaze in trials when the dog chose to gaze at their owner was significantly different by phrase ($F_{3,92} = 4.8$, $p=0.004$, Fig. 3b), but not other factors (dog age: $F_{1,92} = 0.3$, $p=0.583$; Owner sex $F_{1,92} = 0.3$, $p=0.599$, male owner 95% confidence intervals (CI): 1.8–3.5 s vs. female owner CI: 2.1–3.6 s; dog sex $F_{1,92} = 0.7$,

$p=0.417$, male dog CI: 2.1–3.9 s vs. female dog CI: 1.6–3.4 s). The target phrase DDS-meaningful (CI: 3.2–4.7 s) differed from all other phrases: NRP-control (CI: 2.1–3.9 s, $p=0.002$), NRP-meaningful (CI: 1.6–3.3 s, $p=0.014$), and DDS-control (CI: 2.1–3.9 s, $p=0.046$). All other comparisons between DDS-control, NRP-control, and NRP-meaningful were not significant at $p<0.05$ (see ESM for full comparisons).

Fig. 3 **a)** Stacked bar chart plot of binary gaze towards owner by dogs in response to target phrases spoken in: (1) neutral reading prosody (NRP)-control phrase, (2) NRP-meaningful phrase, (3) dog-directed prosody (DDS)-control phrase, and (4) DDS-meaningful phrase. Target phrase had a significant effect on binary gaze response ($F_{3,165} = 12.3$, $p<0.001$), with DDS-meaningful and NRP-Control differing from all other conditions, while NRP-meaningful did not differ to DDS-Control ($p=0.100$). **b)** Box-plot of the duration of gaze (s) in positive reactions to each phrase. Gaze duration of positive reactions did not differ across phrases except for DDS-meaningful to all others ($p<0.05$) but was significant overall ($F_{3,92} = 4.8$, $p=0.004$, $p<0.017$ Bonferroni correction value)



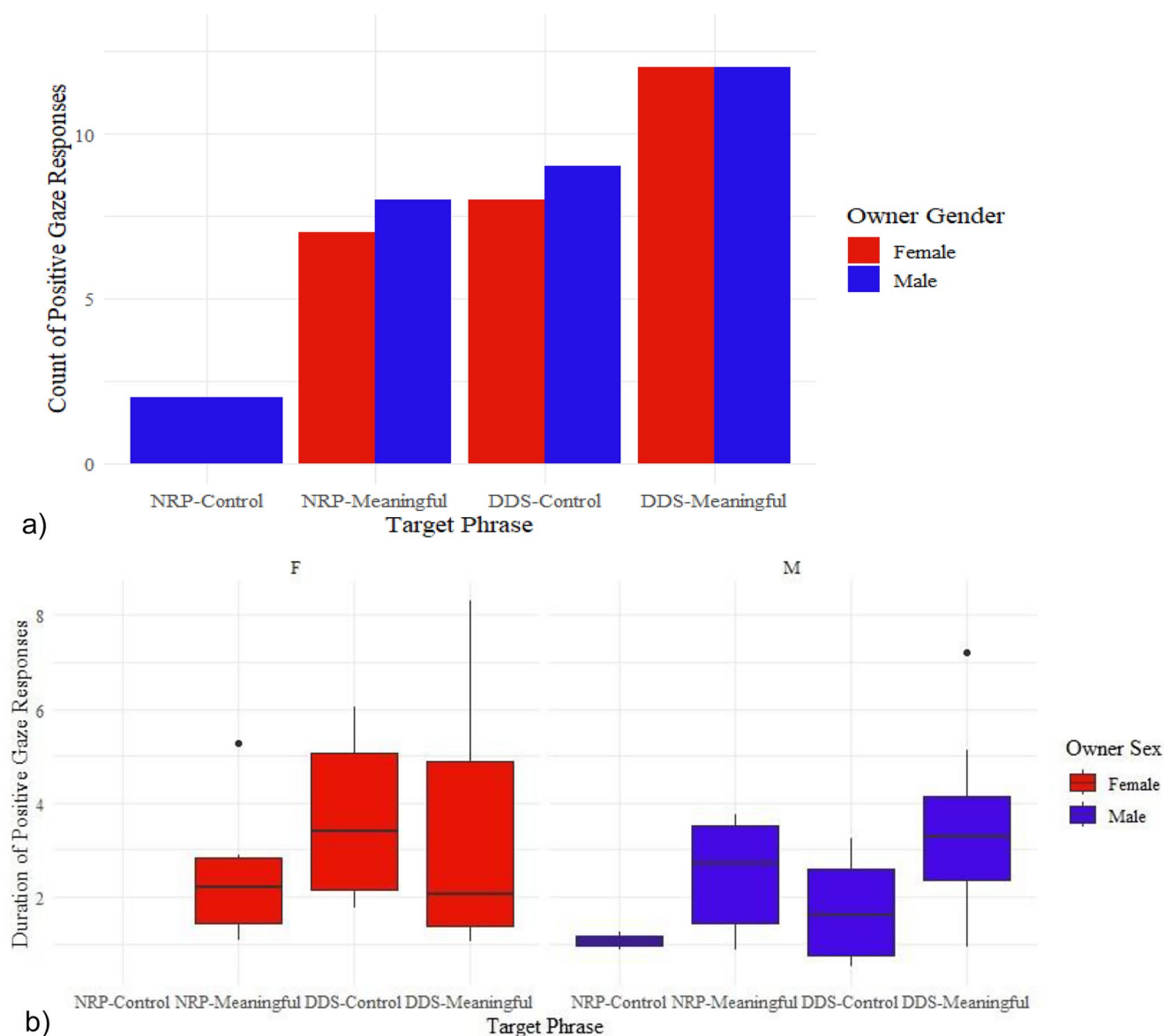


Fig. 4 **a)** Stacked bar chart plot of binary gaze towards owner by dogs in response to target phrases spoken in: (1) neutral reading prosody (NRP)-control phrase, (2) NRP-meaningful phrase, (3) dog-directed prosody (DDS)-control phrase, and (4) DDS-meaningful phrase. Target phrase had a significant effect on binary gaze response ($F_{3,89} = 5.4$, $p = 0.002$), with DDS-meaningful and NRP-Control differing from all

other conditions, while NRP-meaningful did not differ to DDS-Control. **b)** Boxplot of the duration of gaze (s) in positive reactions to each phrase, grouped by owner sex. Gaze duration of positive reactions did not differ across phrases or with owner sex at $p < 0.017$. (Note: No dog responded to NRP-control from their female owner, thus NRP-Control positive gaze duration is empty)

Experiment 2 sex: effects of speaker sex on response to prosody and meaningful content

For Experiment 2 Sex, where we examined the effect of speaker sex on dogs' responses, dogs showed a positive gaze response to their owner in 60.4% of trials (58 of 96, Fig. 4a). The GLMM showed that there was a significant effect of phrase on the binary gaze response ($F_{3,89} = 5.4$, $p = 0.002$), but owner sex ($F_{1,89} = 1.4$, $p = 0.239$), dog age ($F_{1,89} = 1.6$, $p = 0.211$), or dog sex ($F_{1,89} = 0.5$, $p = 0.495$)

had no significant effect. Post-hoc tests showed that all pairwise comparisons of reactions to the four phrase-types were significant at $p < 0.017$ except between NRP-meaningful and DDS-control ($p = 0.517$, see ESM Table 7 for all post-hoc comparisons). Overall, dogs were most likely to look at their owners in the DDS-meaningful condition, and least likely in the NRP-control condition (Fig. 4a).

The LMM showed that duration of gaze in trials when the dog chose to gaze at their owner did not differ significantly

between any fixed effects ($p > 0.3$ for all, see ESM Table 8 for full details, Fig. 4b).

Discussion

We found that dogs spontaneously recognise meaningful phonemic content presented within streams of putatively non-salient speech. Dogs were more likely to look at their owner when the meaningful phrase was presented in dog directed speech (DDS), but they also detected the meaningful phrase when presented in neutral reading prosody (NRP) more often than the control phrase. Thus, dogs' ability to recognise familiar phrase within a stream of speech was not systematically conditional on specific exaggerated prosodic guidance, though this may have been aided by pauses. We support previous results in both dogs (Ratcliffe and Reby 2014; Benjamin and Slocombe 2018) and human infants (Thiessen et al. 2005) which indicated that exaggerated prosodic cues increase attention. Here, we show that while DDS is sufficient to engage dogs' attention without meaningful content, as the dogs responded to the control phrase in DDS at equal rates to the meaningful phrase in NRP speech, content is also important as the dogs responded more often to the combined presentation of DDS with meaningful content and still responded to the presentation of meaningful content presented in NRP. These observations add weight to the idea that dogs separate the speech signals into phonemic and prosodic elements and can respond to both independently (Scheider et al. 2011; Reinholz-Trojan et al. 2012; Andics et al. 2014; Ratcliffe and Reby 2014; Gábor et al. 2020).

When we analysed trials where dogs responded by gazing at their owners, we found that gaze duration was only affected in Experiment 1 and not in Experiment 2. This lack of differences found in Experiment 2 is unlikely to be caused by a ceiling effect as the dogs were given a maximum of 10 s to respond from onset of the included phrase, and no dogs reached this ceiling in Experiments 1 or 2. This may also be a consequence of the small number of dogs that responded to NRP-control presentations (just 4 out of 25 in Experiment 1). Because we were not able to replicate the differences in gaze duration, we consider the finding tentative.

Women produce more exaggerated DDS speech compared to their NRP speech than men. Men also use less DDS prosody when talking to their dog (Prato-Previde et al. 2006), potentially making dogs more familiar with male NRP speech directed at them. Thus, we had predicted that female DDS-speech would be more attention-invoking than male DDS speech, and that male NRP speech would elicit a greater response than female NRP speech. These predictions were not confirmed by our data. Instead, we found that dogs did not respond more readily to presentations of male

NRP or female DDS speech, and sex did not influence dogs' responses to speech content. Indeed, we found no significant effect of sex on the dogs' overall responsiveness or duration of response. Thus, despite differences in how men and women use speech to dogs, dogs showed similar attention to speech prosody across the sexes. We did find a potential effect of dogs' sex on duration of positive gaze response, but the sample size for this was not sufficient to be conclusive and this result should be considered tentative.

In a pet dog's typical environment, most human speech is likely directed towards humans. As such, DDS has been posited to function as a communication strategy aimed at signalling to the dog that the speech is intended for them (Ben-Aderet et al. 2017; Benjamin and Slocombe 2018), and that highly modulated pitch and increased range are attention-getting devices (Lesch et al. 2019). One may thus hypothesise that dogs would require such prosody in order to detect meaningful verbal signals and fail to respond to meaningful content "hidden" (Kaminski et al. 2012) within streams of non-exaggerated speech. However, we found that dogs still have the ability to parse speech in the absence of exaggerated prosodic cueing, an ability that has been hypothesised to be uniquely human and specific to speech perception (Mandel et al. 1995; Reinholz-Trojan et al. 2012). Since speaker voice familiarity has been shown to influence speech perception in human infants (Barker and Newman 2004; Naoi et al. 2012; Trehub 2017), further research could explore whether voice familiarity also affects speech perception in dogs. As it is known that familiar word position within sentences influence infants' ability to segment speech (Seidl and Johnson 2006), testing dogs' responses to a more varied range of target cues, for example embedding their name within a phrase as Floccia et al. (2016) did, would allow direct comparisons with human abilities.

Our behavioural observations lend support to previous indications that non-human animals are capable of segmenting human speech (Hauser et al. 2001; Toro and Trolbalón 2005; Lu and Vicario 2014), including Boros et al.'s (2021) report that dogs possess the neural abilities required for speech segmentation. We suggest that the ability to parse complex vocal utterances is not specific to humans. Furthermore, it may be linked to the ability to recognise alarm call units which can be presented against noise, overlapping calls, or within longer sequences, by conspecifics or heterospecifics. The ability to parse complex call sequences to recognise familiar, salient units is important to vocal learning in general (Elowson et al. 1998; Clay and Zuberbühler 2011) as well as heterospecific call recognition (Morris-Drake et al. 2017; Zhou et al. 2024).

One limitation to our study is the fact that no dogs were present in the room when the dog owners' speech was recorded. This was deliberate in order to avoid the dogs

habituating to the speech. Instead, we asked dog owners to imagine they were talking to their dogs. This less-than-realistic set up may have affected some features of DDS in dog owners' speech (Ben-Aderet et al. 2017). In order to improve the ecological validity of the study (Jeannin et al. 2017) in future experiments dogs could be present during the recording of their owners' DDS. However, we did find the differences in NRP and DDS to be in line with what would be expected for the different speech registers (Burnham et al. 2002).

A further limitation is that speech segmentation studies usually present the speech without including a brief pause before the target phrase. Our speakers produced natural pauses of 0.05–0.6 s, with an average of 0.4s, as a comma was presented before the included phrases. These brief, natural pauses may have helped to cue the dogs to the presence of a new phrase and thus increased their attention to it. However, they were present before all included phrases, as well as at other points in the text, so while they may have increased attention and thus meaningful phrase recognition, we do not believe that they can account for the responses. Further studies specifically presenting the included phrase embedded without a natural pause (and thus coarticulated with surrounding speech) would allow an estimate of the effect of natural pauses in our experiment, as well as further elucidate the extent to which dog speech parsing abilities compare to that of humans.

Our finding is consistent with previous observations that domestic dogs respond to the phonemic content of short speech signals in the absence of exaggerated prosody (Ratcliffe and Reby 2014) and that they use specific brain regions for processing the verbal and nonverbal content of human speech (Andics et al. 2014; Andics and Miklósi 2018; Gábor et al. 2020; Boros et al. 2020). This ability may reflect the effect of selective breeding on dogs to respond to human vocal signals (Hare et al. 2002). To further investigate this, experiments could extend this study to tame wolves (*Canis lupus*), an undomesticated close relative of dogs, as well as to other domesticated species that are regularly exposed to human speech, e.g., cats and horses which have been shown to be sensitive to various aspects of the human voice and speech (Proops et al. 2012; Saito and Shinozuka 2013; Heleski et al. 2015; Galvan and Vonk 2016; Nakamura et al. 2018; Takagi et al. 2019; Saito et al. 2019).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10071-025-01948-z>.

Acknowledgements We thank all the dog owners for their participation in our study. We also thank Jemma Forman, Louise Brown, and Mya Roberts for their invaluable assistance in data collection.

Author contributions HRG, DR, and VR contributed conception and design of the study; HR and AK collected and analysed the data; HRG performed the statistical analysis and wrote the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

Funding The study was funded by the BBSRC under the grant BB/P00170X/1 'How Dogs Hear Us'.

Data availability All coding data files are available as Excel files in the electronic supplementary material and the original videos are available on Dryad (DOI: <https://doi.org/10.5061/dryad.stjq2c1s>).

Declarations

Ethical approval This study was carried out in accordance with the principles of the Basel Declaration and recommendations of the Animal Welfare and approved by the Ethics Review Board (AWERB), University of Sussex ARG/04/04). Approval for recording the human voice was obtained from the University of Sussex's Ethics Review Committee (University of Sussex ERC/33/3). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate All owners were given information and consent sheets and provided written consent to the experiment.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. *Anim Cogn* 10:17–21. <https://doi.org/10.1007/s10071-006-0025-8>
- Andics A, Gácsi M, Faragó T et al (2014) Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Curr Biol* 24:574–578. <https://doi.org/10.1016/j.cub.2014.01.058>
- Andics A, Miklósi Á (2018) Neural processes of vocal social perception: Dog-human comparative fMRI studies. *Neurosci Biobehav Rev* 85:54–64
- Barker BA, Newman RS (2004) Listen to your mother! The role of talker familiarity in infant streaming. *Cognition* 94. <https://doi.org/10.1016/j.cognition.2004.06.001>
- Baru AV (1975) Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog. In: Fant G, Tatham MAA (eds) *Auditory analysis and perception of speech*. ACADEMIC PRESS LTD, London, pp 91–101

- Beerda B, Schilder MBH, Hooff JARAM, De Van, Vries HW (1997) Manifestations of chronic and acute stress in dogs. *Appl Anim Behav Sci* 52:307–319. [https://doi.org/10.1016/S0168-1591\(96\)01131-8](https://doi.org/10.1016/S0168-1591(96)01131-8)
- Ben-Aderet T, Gallego-abenza M, Reby D, Mathevon N (2017) Dog-directed speech: why do we use it and do dogs pay attention to it? *Proc Royal Soc B: Biol Sci* 284:20162429. <https://doi.org/10.1098/rspb.2016.2429>
- Benjamin A, Slocumbe K (2018) ‘Who’s a good boy?!’ Dogs prefer naturalistic dog-directed speech. *Anim Cogn* 1–12. <https://doi.org/10.1007/s10071-018-1172-4>
- Bergelson E, Swingle D (2012) At 6–9 months, human infants know the meanings of many common nouns. *Proc Natl Acad Sci U S A* 109. <https://doi.org/10.1073/pnas.1113380109>
- Binns C, Culling JF (2007) The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am* 122. <https://doi.org/10.1121/1.2751394>
- Bland JM, Altman DG (1997) Statistics notes: Cronbach’s alpha. *BMJ*. <https://doi.org/10.1136/bmj.314.7080.572>
- Boersma P, Weenink D (2009) Praat: doing phonetics by computer. <http://www.praat.org/>
- Boros M, Gábor A, Szabó D et al (2020) Repetition enhancement to voice identities in the dog brain. *Sci Rep* 10. <https://doi.org/10.1038/s41598-020-60395-7>
- Boros M, Magyari L, Török D et al (2021) Neural processes underlying statistical learning for speech segmentation in dogs. *Current Biology* 31:5512–5521.e5. <https://doi.org/10.1016/j.cub.2021.10.017/ASSET/BC8D9189-AEE9-45A3-9B5C-0BEBC46178AE/MAIN.ASSETS/GR2.JPG>
- Bortfeld H, Morgan JL, Golinkoff RM, Rathbun K (2005) Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychol Sci* 16:298–304. <https://doi.org/10.1111/j.0956-7976.2005.01531.x>
- Braem MD, Mills DS (2010) Factors affecting response of dogs to obedience instruction: A field and experimental study. *Appl Anim Behav Sci* 125:47–55. <https://doi.org/10.1016/j.applanim.2010.03.004>
- Brooks JPL, Nooteboom SG (1982) Intonation and the perceptual separation of simultaneous voices. *J Phon* 10. [https://doi.org/10.1016/S0009-4470\(19\)30909-x](https://doi.org/10.1016/S0009-4470(19)30909-x)
- Burnham D, Francis E (1998) Are you my little pussy-cat? Acoustic, phonetic and affective qualities of infant-and pet-directed speech. *Proc 5th International …*
- Burnham D, Kitamura C, Vollmer-Conna U (2002) What’s new, Pussy-cat? On Talking to babies and animals. *Sci* (1979) 296:1435. <https://doi.org/10.1126/science.1069587>
- Cherry EC (1953) Some experiments on the recognition of speech, with one and two ears. *J Acoust Soc Am* 25:975–979. <https://doi.org/10.1121/1.1907229>
- Childers DG, Wu K (1991) Gender recognition from speech. Part II: fine analysis. *J Acoust Soc Am*. <https://doi.org/10.1121/1.401664>
- Clay Z, Zuberbühler K (2011) Bonobos extract meaning from call sequences. *PLoS ONE* 6:e18786. <https://doi.org/10.1371/journal.pone.0018786>
- Deaúx EC, Allen AP, Clarke JA, Charrier I (2016) Concatenation of ‘alert’ and ‘identity’ segments in dingoes’ alarm calls. *Scientific Reports* 2016 6:1 6:1–9. <https://doi.org/10.1038/srep30556>
- Dror S, Miklósi Á, Sommesse A et al (2021) Acquisition and long-term memory of object names in a sample of gifted word learner dogs. *R Soc Open Sci* 8. <https://doi.org/10.1098/RSOS.210976>
- Eloson AM, Snowdon CT, Lazaro-Perea C (1998) Infant babbling in a nonhuman primate: complex vocal sequences with repeated call types. *Behaviour* 135:643–664
- Fairbanks F (1960) The rainbow passage. In: *Voice and Articulation Drillbook*. pp 127
- Fant G, Tatham MAA (1975) Auditory analysis and perception of speech. Elsevier Science
- Fiset S, Beaulieu C, Landry F (2003) Duration of dogs’ (*Canis familiaris*) working memory in search for disappearing objects. *Anim Cogn* 6:1–10. <https://doi.org/10.1007/S10071-002-0157-4/TABL ES/2>
- Fló A, Brusini P, Macagno F et al (2019) Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Dev Sci*. <https://doi.org/10.1111/desc.12802>
- Floccia C, Keren-Portnoy T, DePaolis R et al (2016) British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition* 148:1–9. <https://doi.org/10.1016/j.cognition.2015.12.004>
- Fugazza C, Andics A, Magyari L et al (2021) Rapid learning of object names in dogs. *Scientific Reports* 2021 11:1 11:1–11. <https://doi.org/10.1038/s41598-021-81699-2>
- Fugazza C, Pogány Á, Miklósi Á (2016) Do as I... did! Long-term memory of imitative actions in dogs (*Canis familiaris*). *Anim Cogn* 19:263–269. <https://doi.org/10.1007/S10071-015-0931-8/FIGURES/2>
- Fukuzawa M, Mills DS, Cooper JJ (2005) The effect of human command phonetic characteristics on auditory cognition in dogs (*Canis familiaris*). *J Comp Psychol* 119:117–120. <https://doi.org/10.1037/0735-7036.119.1.117>
- Galvan M, Vonk J (2016) Man’s other best friend: domestic cats (*F. silvestris catus*) and their discrimination of human emotion cues. *Anim Cogn* 19:193–205. <https://doi.org/10.1007/s10071-015-0927-4>
- Gábor A, Gácsi M, Szabó D et al (2020) Multilevel fMRI adaptation for spoken word processing in the awake dog brain. *Sci Rep* 10:11968. <https://doi.org/10.1038/s41598-020-68821-6>
- Gergely A, Faragó T, Galambos Á, Topál J (2017) Differential effects of speech situations on mothers’ and fathers’ infant-directed and dog-directed speech: an acoustic analysis. *Sci Rep* 7:13739. <https://doi.org/10.1038/s41598-017-13883-2>
- Gerken LA (1996) Prosody’s role in Language acquisition and adult parsing. *J Psycholinguist Res* 25:345–356. <https://doi.org/10.1007/BF01708577/METRICS>
- Gibson JM, Scavelli SA, Udell CJ, Udell MAR (2014) Domestic dogs (*Canis lupus familiaris*) are sensitive to the human qualities of vocal commands. *Anim Behav Cogn* 1:281. <https://doi.org/10.12966/abc.08.05.2014>
- Hare B, Brown M, Williamson C, Tomasello M (2002) The domestication of social cognition in dogs. *Sci* (1979). <https://doi.org/10.1126/science.1072702>
- Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78:B53–B64. [https://doi.org/10.1016/S0010-0277\(00\)00132-3](https://doi.org/10.1016/S0010-0277(00)00132-3)
- Heleski C, Wickens C, Minero M et al (2015) Do soothing vocal cues enhance horses’ ability to learn a frightening task? *J Veterinary Behavior: Clin Appl Res* 10:41–47. <https://doi.org/10.1016/j.jvbe.2014.08.009>
- Jeannin S, Gilbert C, Amy M, Leboucher G (2017) Pet-directed speech draws adult dogs’ attention more efficiently than Adult-directed speech. 1–9. <https://doi.org/10.1038/s41598-017-04671-z>
- Johnson KA (2008) Speaker normalization in speech perception. *The Handbook of Speech Perception*
- Jusczyk PW, Houston DM, Newsome M (1999) The beginnings of word segmentation in English-learning infants. *Cogn Psychol* 39:159–207. <https://doi.org/10.1006/cogp.1999.0716>
- Kaminski J, Call J, Fischer J (2004) Word learning in a domestic dog: evidence for fast mapping. *Sci* (1979) 304:1682–1683. <https://doi.org/10.1126/science.1097859>

- Kaminski J, Schulz L, Tomasello M (2012) How dogs know when communication is intended for them. *Dev Sci* 15:222–232. <https://doi.org/10.1111/j.1467-7687.2011.01120.x>
- Kim S, Broersma M, Cho T (2012) The use of prosodic cues in learning new words in an unfamiliar language. *Stud Second Lang Acquis* 34. <https://doi.org/10.1017/S0272263112000137>
- Kuhl PK (1988) Auditory perception and the evolution of speech. *Hum Evol* 3:19–43. <https://doi.org/10.1007/BF02436589>
- Laures JS, Weismer G (1999) The effects of a flattened fundamental frequency on intelligibility at the sentence level. *J Speech Lang Hear Res* 42. <https://doi.org/10.1044/jslhr.4205.1148>
- Lazarowski L, Waggoner P, Hutchings B et al (2021) Maintaining long-term odor memory and detection performance in dogs. *Appl Anim Behav Sci* 238:105301. <https://doi.org/10.1016/J.APPLANIM.2021.105301>
- Lesch R, Kotrschal K, Schöberl I et al (2019) Talking to dogs: companion animal-directed speech in a stress test. *Animals* 9:1–16. <https://doi.org/10.3390/ani9070417>
- Lu K, Vicario DS (2014) Statistical learning of recurring sound patterns encodes auditory objects in Songbird forebrain. *Proc Natl Acad Sci* 111:14553–14558. <https://doi.org/10.1073/PNAS.1412109111>
- Mallikarjun A, Shroads E, Newman RS (2019) The cocktail party effect in the domestic dog (*Canis familiaris*). *Anim Cogn*. <https://doi.org/10.1007/s10071-019-01255-4>
- Mandel DR, Jusczyk PW, Pisoni DB (1995) Infants' recognition of the sound patterns of their own names. *Psychol Sci* 6:314–317. <https://doi.org/10.1111/j.1467-9280.1995.tb00517.x>
- Maros K, Pongrácz P, Bárdos G et al (2008) Dogs can discriminate barks from different situations. *Appl Anim Behav Sci* 114:159–167. <https://doi.org/10.1016/j.applanim.2008.01.022>
- Mattys SL, Jusczyk PW, Luce PA, Morgan JL (1999) Phonotactic and prosodic effects on word segmentation in infants. *Cogn Psychol* 38:465–494
- Matzinger T, Ritt N, Fitch WT (2021) The influence of different prosodic cues on word segmentation. *Front Psychol* 12:622042. <https://doi.org/10.3389/FPSYG.2021.622042/BIBTEX>
- Mazzoni D, Dannenberg R (2015) Audacity [software]
- Mills DS (2005) What's in a word? A review of the attributes of a command affecting the performance of pet dogs. *Anthrozoos* 18:208–221. <https://doi.org/10.2752/089279305785594108>
- Mills DS, Fukuzawa M, Cooper JJ (2005) The effect of emotional content of verbal commands on the response of dogs (*Canis familiaris*). Current issues and research in veterinary behavioral medicine 217–220. <https://doi.org/10.1016/j.applanim.2004.08.025>
- Mitchell RW, Edmonson E (1999) Function of repetitive talk to dogs during play: control, conservation, or planning? *Soc Anim* 7:55–81. <https://doi.org/10.1163/156853099X00167>
- Morris-Drake A, Bracken AM, Kern JM, Radford AN (2017) Anthropogenic noise alters Dwarf mongoose responses to heterospecific alarm calls. *Environ Pollut* 223:476–483. <https://doi.org/10.1016/J.ENVPOL.2017.01.049>
- Nakamura K, Takimoto-Inose A, Hasegawa T (2018) Cross-modal perception of human emotion in domestic horses (*Equus caballus*). *Sci Rep* 8:8660. <https://doi.org/10.1038/s41598-018-26892-6>
- Naoi N, Minagawa-Kawai Y, Kobayashi A et al (2012) Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage* 59. <https://doi.org/10.1016/j.neuroimage.2011.07.093>
- Nygaard LC, Pisoni DB (1998) Talker-specific learning in speech perception. *Percept Psychophys* 60:355–376. <https://doi.org/10.3758/BF03206860>
- Pilley JW (2013) Border collie comprehends sentences containing a prepositional object, verb, and direct object. *Learn Motiv* 44:229–240. <https://doi.org/10.1016/j.lmot.2013.02.003>
- Pilley JW, Reid AK (2011) Border collie comprehends object names as verbal referents. *Behav Process* 86:184–195. <https://doi.org/10.1016/j.beproc.2010.11.007>
- Pongrácz P, Szabo E, Kis A, Andras P, Miklosi A (2014) More than noise? - Field investigations of intraspecific acoustic communication in dogs (*Canis familiaris*). *Appl Anim Behav Sci* 159:62–68. <https://doi.org/10.1016/j.applanim.2014.08.003>
- Prato-Previde E, Fallani G, Valsecchi P (2006) Gender differences in owners interacting with pet dogs: an observational study. *Ethology* 112:64–73. <https://doi.org/10.1111/j.1439-0310.2006.01123.x>
- Prichard A, Cook PF, Spivak M, Chhibber R, Berns GS (2018) Awake fMRI reveals brain regions for novel word detection in dogs. *Front Neurosci* 12. <https://doi.org/10.3389/fnins.2018.00737>
- Proops L, McComb K, Reby D (2012) Cross-modal individual recognition in domestic horses (*Equus caballus*) extends to familiar humans. *Proc Royal Soc B: Biol Sci* 279:3131–3138. <https://doi.org/10.1098/rspb.2012.0626>
- Ratcliffe VF, McComb K, Reby D (2014) Cross-modal discrimination of human gender by domestic dogs. *Anim Behav* 91:127–135. <https://doi.org/10.1016/j.anbehav.2014.03.009>
- Ratcliffe VF, Reby D (2014) Orienting asymmetries in dogs' responses to different communicatory components of human speech. *Curr Biol* 24:2908–2912. <https://doi.org/10.1016/j.cub.2014.10.030>
- Reinholz-Trojan A, Włodarczyk E, Trojan M et al (2012) Hemispheric specialization in domestic dogs (*Canis familiaris*) for processing different types of acoustic stimuli. *Behav Processes* 91:202–205. <https://doi.org/10.1016/j.beproc.2012.07.001>
- Root-Gutteridge H, Ratcliffe VF, Korzeniowska AT, Reby D (2019) Dogs perceive and spontaneously normalize formant-related speaker and vowel differences in human speech sounds. *Biol Lett* 15:20190555. <https://doi.org/10.1098/rsbl.2019.0555>
- Saito A, Shinozuka K (2013) Vocal recognition of owners by domestic cats (*Felis catus*). *Anim Cogn* 16:685–690. <https://doi.org/10.1007/s10071-013-0620-4>
- Saito A, Shinozuka K, Ito Y, Hasegawa T (2019) Domestic cats (*Felis catus*) discriminate their names from other words. *Sci Rep* 9:1–8. <https://doi.org/10.1038/s41598-019-40616-4>
- Scheider L, Grassmann S, Kaminski J, Tomasello M (2011) Domestic dogs use contextual information and tone of voice when following a human pointing gesture. *PLoS ONE* 6:e21676. <https://doi.org/10.1371/journal.pone.0021676>
- Seidl A, Johnson EK (2006) Infant word segmentation revisited: edge alignment facilitates target extraction. *Dev Sci* 9. <https://doi.org/10.1111/j.1467-7687.2006.00534.x>
- Spring M, Polka L, Curtin S (2013) The role of prosody in speech segmentation: comparisons between monolinguals and French-English bilinguals. *Proc Meetings Acoust* 19. https://doi.org/10.1121/1.4801087/16125349/060227_1_ONLINE.PDF
- Takagi S, Arahori M, Chijiwa H et al (2019) Cats match voice and face: cross-modal representation of humans in cats (*Felis catus*). *Anim Cogn*. <https://doi.org/10.1007/s10071-019-01265-2>
- Taylor AM, Ratcliffe VF, McComb K, Reby D (2014) Auditory communication in domestic dogs: vocal signalling in the extended social environment of a companion animal. Elsevier
- Taylor AM, Reby D, McComb K (2009) Context-related variation in the vocal growling behaviour of the domestic dog (*Canis familiaris*). *Ethology* 115:905–915. <https://doi.org/10.1111/j.1439-0310.2009.01681.x>
- Thalmann O, Shapiro B, Cui P et al (2013) Complete mitochondrial genomes of ancient Canids suggest a European origin of domestic dogs. *Sci* (1979) 342:871–874. <https://doi.org/10.1126/science.1243650>
- Thiessen ED, Hill EA, Saffran JR (2005) Infant-Directed speech facilitates word segmentation. *Infancy* 7:53–71. https://doi.org/10.1207/s15327078in0701_5

- Toro JM, Trobalón JB (2005) Statistical computations over a speech stream in a rodent. *Percept Psychophys* 67:867–875. <https://doi.org/10.3758/BF03193539/METRICS>
- Trehub SE (2017) The maternal voice as a special signal for infants. Early vocal contact and preterm infant brain development. *Bridging the Gaps Between Research and Practice*
- Valian V, Levitt A (1996) Prosody and adults' learning of syntactic structure. *J Mem Lang* 35:497–516. <https://doi.org/10.1006/JMLA.1996.0027>
- Vihman MM (2017) Learning words and learning sounds: advances in Language development. *Br J Psychol* 108. <https://doi.org/10.1111/bjop.12207>
- Vilà C, Savolainen P, Maldonado JE et al (1997) Multiple and ancient origins of the domestic dog. *Science* (1979) 276:1687–1689. <https://doi.org/10.1126/science.276.5319.1687>
- Volsche S, Root-Gutteridge H, Korzeniowska AT, Horowitz A (2023) Centring individual animals to improve research and citation practices. *Biol Rev* 98. <https://doi.org/10.1111/brv.12912>
- Westermann G, Mani N (2017) Early word learning
- Xu N, Burnham D, Kitamura C, Vollmer-Conna U (2013) Vowel hyperarticulation in parrot-, dog- and infant- directed speech. *Anthrozoos* 26:373–380. <https://doi.org/10.2752/175303713X13697429463592>
- Yin S, McCowan B (2004) Barking in domestic dogs: context specificity and individual identification. *Anim Behav* 68:343–355. <https://doi.org/10.1016/j.anbehav.2003.07.016>
- Zhou Y, Radford AN, Magrath RD (2024) Noise constrains hetero-specific eavesdropping more than conspecific reception of alarm calls. *Biol Lett* 20. <https://doi.org/10.1098/RSBL.2023.0410>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.