scientific reports

OPEN

Sound symbolism facilitates interspecies communication between humans and domestic dogs *(Canis familiaris)*

A. T. Korzeniowska[1,2], H. Root-Gutteridge[3], D. Reby[4] & J. Simner[1]✉

The evolution of human communication likely centred, in part, on shared intuitions about the mapping of sound to meaning. These sound-meaning intuitions, known as *sound symbolism*, can be seen for example in the *bouba-kiki* effect, where nonsense words carry inherent meaning about their likely referents (here, rounded vs. angular objects respectively). In our paper we suggest for the first time that sound symbolism can afford successful interspecies communication between humans and animals in certain circumstances. Over four investigations, including replications, we show that humans use sound symbolism significantly and pervasively to attempt to convey meaning to domestic dogs (*Canis familiaris*), specifically, by exploiting vocal prosody to signal elevation in space. In Study 1 we analysed recordings of amateur dog owners commanding their dogs to move upwards (e.g., "stand") or downwards (e.g., "down"), finding higher mean pitch (fundamental frequency, f_0) in the former versus the latter. In Studies 2 and 3a, we replicated this in competitive dog owners, both in self-report, and in acoustic voice-analyses recorded in competition. In addition, professionals also used further sound symbolism beyond amateurs, in their commands for the dog to "sit" (using higher pitch to denote sit up vs. sit down). Finally, in Study 3b, we demonstrate that sound symbolism appears to be mutually understood by dogs in certain useful circumstances. Dogs were faster to enact "down" commands with prosodic sound symbolism, compared to without, demonstrating that sound symbolism may sometimes underlie successful inter-species communication.

Keywords Dog, Sound symbolism, Prosody, Cross-modality, Pitch, Animal communication

Our studies centre on the question of whether humans and dogs can establish mutual understanding using *sound symbolism*. Human utterances have sound symbolism if they directly express meaning from their acoustic properties (for review see[1]. In languages with relatively well-developed sound symbolism, even non-native speakers can guess word-meanings. For example, monolingual English speakers can guess the meanings of Japanese foreign dimensional adjectives (e.g., "big" vs. "small") at above-chance level (also in Korean, Albanian, Gujarati, and so on; e.g[2]. , suggesting that meaning is in some way carried by the way words sound. English itself has comparatively little sound symbolism, although rare instances might still be sufficient to establish communication with animals. We describe the relevant sound symbolism of English below.

English has sound symbolism in a small class of onomatopoeic words (e.g., "buzz") but also, for example, in the tendency to link high-front vowels (/I/) with small objects, low back vowels (/a/) with large objects[3,4], and rounded vowels with rounded objects (this latter demonstrated by the *bouba-kiki* effect;[5,6]. Recent studies had suggested that only humans understand sound symbolic language; Bonobos (*Pan paniscus*), for example, show no evidence of the *bouba-kiki* effect, despite direct testing[7]. But sound symbolism can also be found in English prosody, which comprises the pitch, intonation, rhythm or loudness of speech (expressed acoustically with features such as the mean fundamental frequency (f_0), variation in f_0, duration, and amplitude). Although prosody was once considered referentially redundant, Nygaard et al.[8] have argued that certain semantic domains bear unique acoustic signatures. One example – relevant here – is pitch variation when referring to objects in the vertical plane. In testing conditions, humans spontaneously speak with higher pitch when referring to objects moving upwards compared to downwards[9]. This type of study provides empirical support for a link between

[1]School of Psychology, University of Sussex, Falmer, Brighton, UK. [2]eLife Sciences Publications Ltd, Cambridge, UK. [3]School of Life and Environmental Sciences, University of Lincoln, Lincoln, UK. [4]Equipe de Neuro-Ethologie Sensorielle Bioacoustics Research Laboratory, Centre de Recherche en Neurosciences de Lyon, University of Saint-Etienne, CNRS, Inserm, Saint-Etienne 42100, France. ✉email: J.Simner@sussex.ac.uk

prosody and meaning (i.e., vertical position in space). Since prosody is the basis for our studies we expand on its sound symbolic properties below.

The referential use of prosody (e.g., higher pitch in the human voice [f_0] intuitively denotes higher in space) also aligns with *crossmodal correspondences*, which are another set of human intuitions but often at a more sensory level (e.g., higher sinusoidal pitch intuitively denotes higher in space; for review see[10]. Hence, correspondences align with prosodic sound symbolism because humans pair high elevation with high pitch, whether from pure tones (i.e., a correspondence;[11] or linguistic prosody (i.e., sound symbolism;[9]. Also important is that crossmodal correspondences exist in non-human species, including chimpanzees (*Pan troglodytes*;[12], domestic dogs[13], chicks (*Gallus domesticus*;[14] and tortoises (*Testudo hermanni*[15]. For example, we have shown that dogs engage more with audio–visual stimuli that are congruent with human pitch-elevation correspondences (i.e., they look longer at spatially-elevated objects making higher vs. lower pitch noise;[13]. This suggests crossmodal correspondences may be a broad feature of neural systems, and not uniquely human. In the current paper we therefore explore whether this type of crossmodal correspondence (linking acoustic pitch and elevation) may play a role in the linguistic communication between humans and domestic dogs, specifically in the context of prosodic sound symbolism. Across four investigations, we first explore whether humans use vocal prosody to attempt to convey meaning to dogs (using higher pitch speech to denote objects higher in space). Perhaps most importantly, we explore whether this sound symbolic speech of dog owners is ever understood by dogs themselves.

In Study 1, dog owners gave six commands to their dogs: two instances of "stand", "down", and "sit" (without hand signals). "Stand" instructs the dog to be fully upright, touching the ground with four paws; "sit" is a mid-position with rump on the ground but up on front paws; and "down" is a fully low position when the dog's rump, stomach and elbows are touching the ground. Table 1 shows the movement in the vertical plane associated with these six commands, depending on the dog's starting position. We hypothesised that owners may use a higher spoken pitch (mean f_0) when commanding dogs to move upwards, compared to downwards. Here we compared either *across* lexemes (comparing "stand" to "down"), and *within* a lexeme (comparing "sit" when body moves up [from a lying position] to "sit" when body moves down [from a stand position]; henceforth we call this "sit"-up and "sit"-down). We explored whether one or both comparisons may show prosodic sound symbolism (i.e., higher mean f_0 when commanding the dog to move upwards vs. downwards).

As well as acoustic data, we captured questionnaire information from owners about their dog-owning experience (in months), training experience (from "none" to "professional"), dog's overall training level (informal, formal, sport training), and dog's proficiency (in "sit", "down", "stand"). We hypothesised that greater experience or proficiency may increase the likelihood of an owner using sound symbolism. Note that even if sound symbolism were universal and/or innate, we still expect a training effect because people are self-aware. (A useful analogy is smiling, which is instinctive and cross-cultural, but people can still learn to smile more or even artificially in certain situations, simply because they understand its communicative value.)

In a second study we recruited and tested more experienced dog owners, who train for the Competition Obedience sport of Distant Control (as regulated by the UK Kennel Club[16] where dogs are given the same six commands as in Table 1, but in an arena at a distance of 10–20 paces. Sound symbolic prosody may be particularly beneficial here because dogs must respond quickly and accurately in a difficult communicative context (i.e., in a noisy arena from afar). Hence in Study 2 we asked, via questionnaire, how competitive dog handlers uttered "stand" and "down" commands, in terms of high (or rising) pitch versus low (or falling) pitch (henceforth, high or low, respectively). (Note that we additionally examine "sit"-up and "sit"- down in Study 3.) We again predict higher pitch when commanding the dog upwards versus downwards, and again elicited handler's experience (i.e., number of dogs trained in Competition Obedience, and Kennel Club competition class achieved).

In a final pair of studies (3a/b) we asked the same questions but in a naturalistic competition setting, conducting acoustic analyses of dog-handlers' commands (Study 3a) and observing the subsequent behaviour of their dogs (Study 3b). Here we scrutinised video records of the Championship Obedience Distant Control competitions at Crufts, a pre-eminent international dog show hosted annually by the United Kingdom Kennel Club, and the largest competition of its kind in the world. For acoustic analyses, we measured hander's mean f_0 in their commands for "stand", "down" and "sit" (both "sit"-up, and "sit"-down). As before, we hypothesized handlers would recruit the sound symbolism (i.e., higher f_0 for commands to move upward vs. downwards). Crucially, we also hypothesised an impact on dogs themselves: dogs whose handlers employed greatest referential prosody (lower pitch = down; higher pitch = up) should react fastest. This pattern of results could provide evidence of

Command uttered	Starting position (before command)	Movement required
"down"	**stand** position	↓
	sit position	↓
"stand"	**down** position	↑
	sit position	↑
"sit"	**stand** position	↓
	down position	↑

Table 1. Six obedience commands used across our studies. Arrows indicate the direction of the required movement (up or down) in vertical space, given the starting position. Since "sit" can denote moving either upwards or downwards (depending on starting position) we henceforth describe sit-from-lying as "sit"-up and sit-from-standing as "sit"-down.

successful interspecies communication based on sound symbolism. Specifically, we predicted that dogs would respond faster to sound symbolism in "stand" and "down" commands, but not in "sit" commands, because pitch is redundant when commanding a dog to sit. Hence even if it is uttered, it cannot help the dog. This is due to the fact that "sit" is uttered in contexts where there is only a single possible direction of movement -- either when the dog is standing (only movement is downwards) or when the dog is lying (only movement is upwards). In contrast, commands like "down" and "stand" can be given while the dog is seated, a position from which two directions are possible (up to standing or down to lying). Thus, while prosody can guide interpretation for "down" and "stand" commands, it offers no useful information for "sit" commands. We therefore predict dogs may respond to sound symbolism in "down" or "stand", but not where it is referentially redundant in "sit" commands.

Results

Study 1: acoustic analysis of prosody in voice commands from non-competitor dog owners

Our participants were 34 dog owners (31 female, 2 males, 1 another gender; median age fell within 36–45 years) who had owned dogs for an average of 174.03 months (SD = 188.68). The mean age of their current dog was 55.68 months (SD = 36.35). Twenty-two owners had trained their dogs in basic obedience only, 6 in dog sport (but only 1 competitively), 5 were dog trainers, and 1 had no dog training experience. Here and throughout, participants were unpaid volunteers. Each participant was asked to submit a recording of 6 commands (two of "stand", two of "down", and two of "sit"; see Table 1). Participants submitted 186 useable recordings (with a further 18 missing or unusable; e.g., from background noise; more details can be found in the raw datafile on the OSF). We ran an acoustic analysis of their speech when commanding their dogs in "stand", "down" and "sit".

Across all studies, we compared the pitch of commands either across lexemes ("stand" vs. "down") or within a lexeme ("sit"-up vs. "sit"-down). In both comparisons, the dog's body was moving upwards in the former token and downwards in the latter (see Table 1). We predict a higher mean f_0 in commands for the dog to move upwards ("stand" or "sit"-up) compared to downwards ("down" or "sit"-down). We analysed "sit" away from "stand"/ "down" for four important reasons. First, the direction of movement for "sit" depends on context, while in "stand"/"down" it does not; i.e., only one direction is signalled for each of the latter ("stand" ↑; "down" ↓), while two are possible for the former ("sit" ↑↓), because the direction of movement depends on starting position (see Table 1). Second, analysing "sit" in isolation provides a useful within-lexeme comparison (i.e., when the dog is commanded to sit up or sit down, there is no change in the segmental pronunciation – "sit" -- but meaning still changes). This allows us to understand the precise calibration of any sound symbolism: i.e., does sound symbolism arise only across different lexemes ("stand" vs. "down"), or does it even arise within a single lexeme ("sit", meaning up or down). The third reason we analysed "sit" away from "stand"/ "down" is due to a difference in the manner of body's elevation. The "stand" and "down" commands result in the dog's body being fully elevated or fully ground-level, while the "sit" command results in the body being partially elevated (either because only the head is raised from lying, or only the rump is lowered from standing). The fourth and most important reason we analysed "sit" away from "stand"/ "down" is to avoid a confound, because "sit" has a notably different vowel quality to "stand" vs. "down". Compare the initial low vowel in /stand/ and /daʊn/ to the vowel in /sIt/, which is inherently higher in pitch[17]. Not only is pitch one of our dependent measures (therefore important to not conflate across items), but the vowel in "sit" may even have its own set of sound symbolic properties, which are unrelated to our focus here; e.g., its high-front vowel (/I/) is also associated with small objects[3,4]. Hence for all four reasons, we analysed "sit" away from "stand"/ "down", here and throughout.

All analyses were conducted using R Studio, R version 4.1.2[18]. For evidence of sound symbolism, we employed Linear Mixed Models (using *lme4* package[19], with mean f_0 as outcome, type of command as a predictor (Model 1: "stand" vs. "down", Model 2: "sit"-up vs. "sit"-down), and Handler ID as a random effect (both intercepts and slopes). Our data showed some evidence of sound symbolism. Our Linear Mixed Models revealed significantly higher mean f_0 in "stand" (279 Hz, SE = 13.19) versus "down" commands (208 Hz, SE = 9.61), as hypothesised ($F(1,29.83) = 36.97$, $p < 0.0001$). However, there was no significant difference in pitch between "sit"-up (353 Hz; SE = 21.0) and "sit"-down (346 Hz, SE = 21.4; $F(1, 30.37) = 0.19$, $p = 0.67$; see Fig. 1a/b respectively).

We also compared the degree of sound symbolism against the life-experience of dog and owner. Specifically, we compared the size of the difference in mean f_0 for "down" versus "stand" (and again for "sit"-up vs. "sit"-down) against: (a) length-of-owning-dogs (in months, using Spearman correlation within the *ggpubr* package[20] (b) owner's training experience, splitting owners into high experience (training for sport or competition, including professionals) versus low experience (no dog training or basic obedience) (c) dog's training experience (informal, formal, sport training) (d) dog's proficiency with the question: *Does your dog react to a vocal cue for "sit"/"down"/"stand" without needing a hand signal?*; answers were: *Yes* (coded 3), *Most of the time* (coded 2), *Sometimes* (coded 1), and *No* or *My dog was not trained to sit/ go down/ stand on cue* (coded 0). For variables (b)-(d) we predicted mean f_0 difference using linear models (*stats* package *lm* and *aov* function from the *stats* package[18]. None of our effects were significantly modulated by the dog-owner's experience (stand/down r_s=0.17, $p = 0.37$; "sit"-up/down r_s=0.25, $p = 0.17$), nor the dog owner's training experience (stand/down b=-28.98, $t(29)$=-1.29, $p = 0.21$; "sit" -up/down b = 5.10, $t(29) = 0.21$, $p = 0.83$), nor the dog's proficiency for commands (stand/down b = 18.51, $t(29) = 1.85$, $p = 0.07$; sit up/down b = 12.53, $t(29) = 0.49$, $p = 0.63$), nor the dog's overall training (stand/down ($F(2,10.58) = 2.77$, $p = 0.11$; sit-up/down ($F(2,28) = 0.2$, $p = 0.82$).

Study 2: Self-report of prosody in voice commands from competitors

We tested 128 competition dog owners (8 males, 119 females, 1 undisclosed; median age fell within 56–65 years) who trained their dogs in the Competition Obedience Distant Control exercise as regulated by the UK Kennel Club[16]. Participants completed questionnaires describing their prosody (high or low pitch) for the commands of interest (i.e., "stand" and/or "down") and provided 240 valid responses (at least one per participant). Our

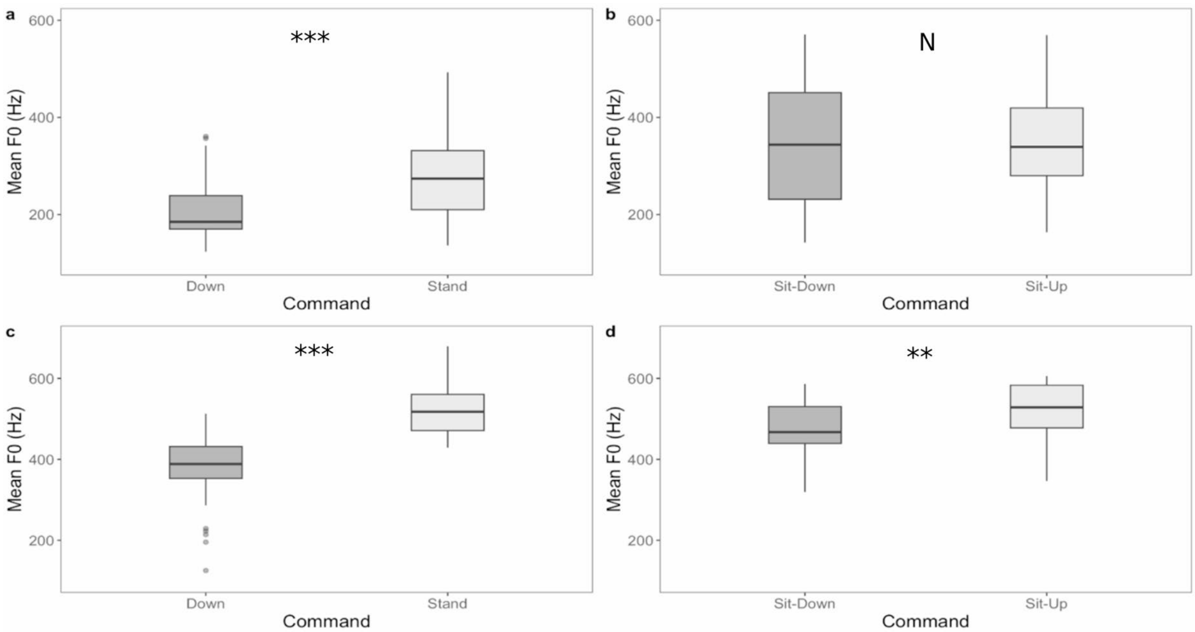

Fig. 1. Mean f_0 in commands given by owners to their dogs. Amateur dog owners showed (**a**) significantly higher mean f_0 in "stand" versus "down", and (**b**) no difference in f_0 for "sit"-up versus "sit"-down. Competition dog owners in Crufts Championship Obedience Competition showed (**c**) a significantly higher mean f_0 in "stand" versus "down", as well as (**d**) in "sit"-up versus "sit"-down. In these plots, the box represents the interquartile range, the line within the box indicates the median, the whiskers extend to the smallest and largest values within 1.5 times the interquartile range, and the individual points outside the whiskers are outliers.

target population for analyses were those who reported manipulating pitch when commanding the dog i.e., 132 out of the 240 valid responses. The remainder used loudness (N45); no voice modulation (N22); no voice at all (hand signal only; N18); or other (N23). We additionally tested but later removed 43 further participants with incomplete questionnaires (e.g., no information on how they met Kennel Club regulations; see below).

We analysed our participants separately as two groups because Kennel Club regulations[16] allow handlers to either use a vocal command (e.g., "stand") or a body signal (e.g., hand moves up), but not both – and either is optionally preceded by the dog's name. Our target groups were therefore: (a) the 'body signal' group: here we analysed self-reported pitch in the dog's name (i.e., the only word uttered; 31 responses comprising 23% of targets) or (b) 'vocal command' group: here we analysed self-reported pitch in the vocal commands "stand"/"down" (which perfectly matched their report of the pitch of any dog name used); these were the remaining 101 responses comprising 77% of targets. Importantly, whether analysing dog name or command, we broke down by command-type (i.e., dog name or command for the instruction to stand vs. lie down).

We predicted pitch-elevation sound symbolism, which would show itself as more high pitch for "stand" commands, and more low pitch for "down" commands. For this we employed a binomial GEE for each target group (using the *geeglm* function from the *geepack* package[21], with a binary output of low/high pitch, and a predictor of whether the command was stand or down (accounting for repeated measures, because handlers could report pitch for both commands).

The 'vocal command' group showed sound symbolism (binomial GEE, $b = 4.39$ ($SE = 0.7$), Wald χ^2 $(1) = 39.7$, $p < 0.0001$), with more high pitch prosody reported for stand commands (28 participants; vs. 6 who reported using low pitch), and more low pitch prosody reported for down commands (63 participants; vs. 4 who reported using high pitch). This significant pattern was repeated in the 'body signal' group (binomial GEE, $b = 36.0$ ($SE = 0.47$), Wald χ^2 $(1) = 5924$, $p < 0.0001$), with more high pitch prosody reported for stand commands (17 participants; vs. 6 who used low pitch), and more low pitch prosody reported for down commands (8 participants; vs. 0 who used high pitch). This replicated our findings for "stand"/ "down" in Study 1, but this time for professional dog handlers.

We again asked whether any effects were linked to the handler's experience (model 1 = number of dogs trained; model 2 = highest obedience class achieved). We ran binary logistic regressions (using the *stats* package *glm* function[18] with a target population who used pitch for *both* commands ("stand" and "down"). The binary outcome was Yes/No for whether prosody was congruent with the pitch-elevation correspondence (congruent = higher pitch for stand commands, and lower pitch for down commands). These analyses were run for the 'vocal command' group only, given small numbers in the 'body signal' group ($N = 6$). As in Study 1, our effects were again unrelated to handler experience. Specifically, our binary logistic GLM model first re-confirmed the above pattern of sound symbolism, for both the 'vocal command' group ($b = 1.39$, $SE = 0.46$, $p = 0.002$) and 'body signal' group (here, all participants who used pitch in both "down" and "stand" command [N6] did so in agreement with the pitch-elevation correspondence). And where there were sufficient numbers (in the 'vocal command' group), we found no effect of handler's competition class (b = 0.05, SE = 0.27, $p = 0.87$) nor number of dogs trained (b=-1.9, SE = 2.22, $p = 0.39$).

Study 3a: acoustic analysis of prosody in voice commands from competitors

Here we ran an acoustic analysis of the speech of dog handlers, when commanding their dogs in competition. We identified 76 target videos of the Crufts Distant Control competition containing (after exclusions, see SI) 201 voice files from 35 handlers, comprising 33 females and 2 males. These included 27 tokens of "stand", 125 tokens of "down", 24 tokens of "sit"-up, and 25 tokens of "sit"-down (with varying numbers of commands per handler depending on how many dogs they competed with, and how many of their commands were excluded from the analysis; see raw data file on the OSF for details). Analyses and predictions were as in Study 1. As before we found significantly higher mean pitch (f_0) in "stand" (522 Hz, SE = 25.5) versus "down" commands (385 Hz; SE = 12.3), as hypothesised ($F(1,7.84) = 22.23$, $p = 0.002$). Unlike for amateur owners (Study 1), there was also significant difference between "sit"-up (537 Hz, SE = 26.2) and "sit"-down (489 Hz; SE = 17.2; $F(1,11.94) = 6.66$, $p = 0.02$; see Fig. 1c/d), with the former higher in pitch.

We found an almost identical pattern of results in our supplementary analyses, when we included commands that had been preceded by the dog-name. Again we found significantly higher f_0 in "stand" (482 Hz; SE = 18.6) versus "down" commands (384 Hz; SE = 11.7) with a similar effect size as before ($F(1,29.80) = 28.10$, $p < 0.0001$). Similarly, "sit"-up commands, too, were uttered with a higher pitch (523 Hz; SE = 23.5) than the "sit"-down commands (483 Hz; SE = 18.9), with the effect approaching significance ($F(1,16.74) = 3.40$, $p = 0.08$).

Study 3b: dogs' performance as a function of sound symbolic prosody in voice commands

Here we analysed how dogs reacted to these same commands ("down", "stand", "sit"-up, "sit"-down) in this same competition (Crufts Distance Control). We analysed whether mean f_0 in the handler's voice command (taken from Study 3a) might have affected the dogs' performance (coded in Study 3b). We analysed 68 target videos, involving 30 handlers and 44 dogs (some handlers competed with multiple dogs over the years). The number of responses per dog again varied depending on how many times across the years they competed, and how many of their response videos were excluded (see data file on OSF for details). Dogs had a mean age 7.4 years (SD = 2.8); half were female and all were Border Collies, or Working Sheepdogs (i.e., non-pedigree Border Collies, who may or may not work sheep).

As before, our primary analyses were conducted on the "vocal command group" (who did not use hand signals), including only those who did not use the dogs' name. As in Study 3a, we also ran supplementary analyses including 56 additional tokens from handlers who gave vocal commands preceded by the dog-name. Here, we did not splice the command from the dog's name but extracted mean pitch (f_0) from the entire utterance (e.g., "Buster, stand"). This is because one of our measurements, the dog's latency to react, was likely to happen from the very start of the utterance (i.e., moment of the dog's name), and not further downstream at the command.

For each type of command ("stand", "down", "sit"-up, and "sit"-down) we ran two Linear Mixed Models with mean f_0 as a predictor, and Dog ID as a random factor. The outcome variables were latency-to-move (Model 1), and duration-of-movement (Model 2). For "down" commands we predicted faster latencies and durations-of-movement for lower pitch utterances (and the reverse for "stand"; i.e., faster for higher pitch). Such results would suggest prosodic sound symbolism is understood by dogs, and aids comprehension. For "sit" commands, we predicted no effect of pitch on the dog's behaviour, because pitch is non-informative for "sit" commands (see above and Discussion).

As predicted, mean pitch did not affect "sit" commands: i.e., no significant effect of pitch on latency-to-move nor duration-of-movement, either for "sit"-up commands ($F(1,18.51) = 0.11$, $p = 0.74$; $F(1,17.64) = 0.03$, $p = 0.86$, respectively) nor "sit"-down commands ($F(1,4.91) = 4.10$, $p = 0.10$; $F(1,19) = 0.80$, $p = 0.38$, respectively). There was no effect either when we added in commands with dog-names ("sit"-up latency-to-move $F(1,23) = 0.0002$, $p = 0.98$, and duration-of-movement, $F(1,21.35) = 0.16$, $p = 0.70$; "sit"-down latency-to-move $F(1,20.26) = 2.60$, $p = 0.12$, and duration-of-movement $F(1,23) = 0.18$, $p = 0.68$).

We did however find an effect in "down" commands, as predicted. Here, although pitch did not affect the dogs' latency-to-move (for our main analysis $F(1,73.60) = 0.40$, $p = 0.53$, nor when incorporating additional commands including the dog's name $F(1,77.80) = 0.30$, $p = 0.58$), it did significantly predict their duration-of-movement. As such, dogs moved into position faster if the mean pitch of "down" was lower ($F(1,77.35) = 5.71$, $p = 0.019$; see Fig. 2). We found an identical pattern of results when we added in data from handlers who also used their dog's name ($F(1,82.01) = 6.14$, $p = 0.015$). In summary, our finding shows that, across handlers/dogs, the dogs who moved down fastest were those whose handlers employed low pitch.

Finally, contrary to predictions, mean pitch had no effect in "stand" commands, neither for latency-to-move (main analysis: $F(1,6.14) = 0.52$, $p = 0.50$; supplementary analysis: $F(1,50.12) = 0.30$, $p = 0.60$) nor duration-of-movement (main analysis: $F(1,16.52) = 0.002$, $p = 0.97$; supplementary analysis: $F(1,63.98) = 0.05$, $p = 0.82$).

Discussion

Our data suggest humans employ prosodic sound symbolism when speaking with their dogs, and that in certain circumstances dogs appear to show reciprocal understanding. Owners used higher pitch to urge their dog to move upwards ("stand") versus downwards ("down"), matching previous findings of sound symbolic prosody in human-to-human interactions[9]. We found prosodic sound symbolism for "stand/down" commands in both self-report and acoustic analyses, in both experimental and naturalistic settings, and in both amateur and competition-trained dog owners, all the way up to Crufts Championship competitors. This initially suggests it may be a tendency unrelated to dog-handling experience. However, only competition handlers used prosodic sound symbolism to distinguish by pitch two meanings from a single lexeme ("sit"→ sit up, sit down), while amateur dog owners did not. One explanation is that amateurs may not realise a single word "sit" can have

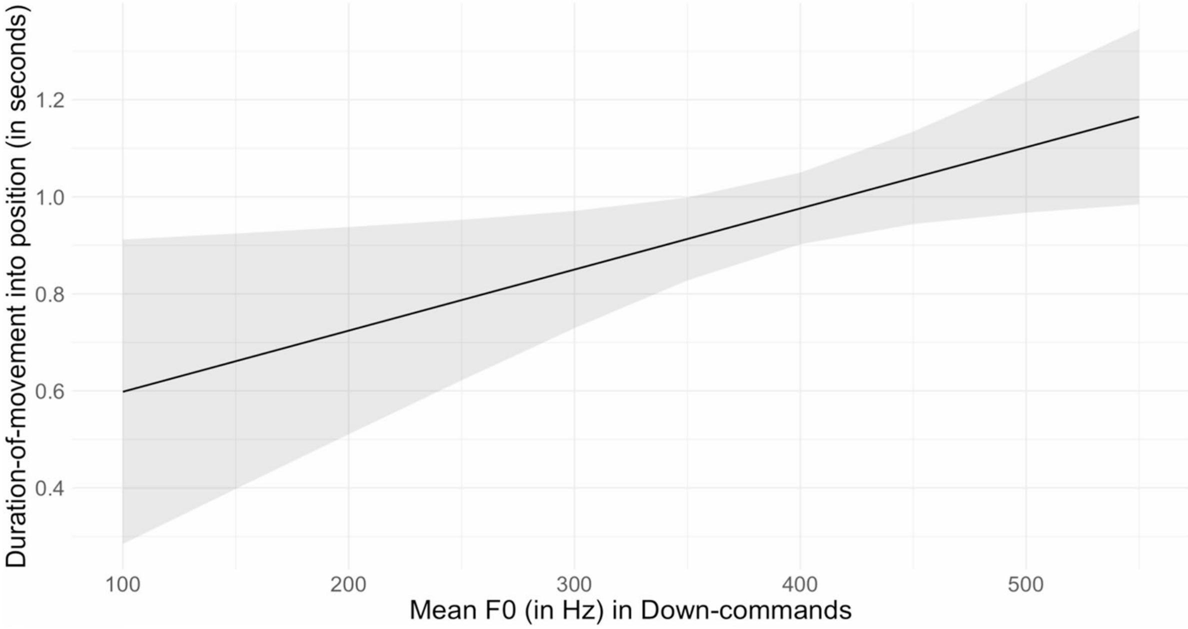

Fig. 2. Duration-of-movement predicted from mean f_0 in "down" commands. Dogs were faster to move into a down position as mean pitch lowered. Shading represents 95% confidence intervals.

different directional meanings (i.e., they perhaps treat "sit" as singular command to mean "rest on rump/front paws", without considering direction of movement). An alternative interpretation is they implicitly choose not to employ pitch because pitch is referentially redundant when applied to "sit". As described above, the "sit" command is always given in a context where only one direction of movement is possible so prosody cannot help (see below where we discuss this further). In contrast, competition handlers *still* apply pitch even in "sit" commands, despite its redundancy, perhaps as part of a broader training strategy. Competition handlers may feel (consciously or otherwise) that consistently training dogs to associate pitch with directionality will help dogs apply that prosodic cue in situations where it *is* informative—such as with "down" or "stand" commands. In fact, redundant pitch-elevation prosody is also found elsewhere, in studies between humans[9], and we observe a similar phenomenon here.

Not only did we demonstrate sound symbolism when humans speak to dogs, we also showed some evidence of mutual understanding. In Study 3b, pitch modulation predicted dogs' behaviour because dogs were faster to move downwards when "down" was uttered in a lower (vs. higher) pitch. This provides the first indication of successful inter-species communication based on sound symbolism in human language. We can be confident dogs were responding to voice, because bodily cues (e.g., hand or head movements) were specifically prohibited by competition regulations in the data we analysed (and any theoretical eye-movements mimicking the dogs' movement downwards would be only in the region of 1-2º so could not be detected by dogs at the competition distance of around 33 feet). Furthermore, we are confident dogs could perceive the pitch changes in speech prosody because they are capable of distinguishing auditory frequencies to a satisfactory degree[22], and their hearing range is similar to humans[23,24]. This means dogs can hear prosodic pitch changes and potentially find them meaningful, as our data for "down" suggest. Our findings marry with animal studies that have previously shown dogs linking artificial stimuli (e.g., matching machine-generated noise to the size or position of artificial objects; e.g[13,25,26]. But no previous study has shown sound symbolism in verbal utterances giving some degree of inter-species understanding, as we've shown here. Our data also follow a long line of anecdotal reports that different trainers use similar non-linguistic sounds across various cultures (see[26] (e.g., whistles, clicks, claps; e.g[27–31]), but again, there had been no previous evidence as to whether these human-produced trainer-signals were effective.

We did not predict, and did not find, dogs responding to the prosody of "sit". Their competition handlers used higher pitch when "sit" meant move upwards (from lying) versus downwards (from standing), but prosody did not impact the dogs' behaviour. We noted above that prosody here is non-informative because, unlike "down" or "stand" commands, "sit" is only given in contexts where prosody cannot help. Hence if the "sit" is instructed from standing, it means move downwards, but the only other possible command from that position also means move downwards (i.e., "down"). Equally, if "sit" is instructed from lying, it means move upwards, but the only other possible command from that position also means move upwards (i.e., "stand"). For this reason, prosody is not a useful feature when uttering "sit" (because context means there is only ever one possible direction of movement). In contrast, "down" commands can be given while the dog is seated, where two directions are theoretically possible (up to standing or down to lying) so pitch can help clarify the intended movement. The same applies to "stand" commands (because they too can be given from sitting). Hence prosody is helpful for "down/stand" commands in a way it is not helpful for "sit". Of course, our dogs' null effect for "sit" may alternatively be from lack of power, but this redundancy means prosody modulation was specifically not predicted to influence the dogs' movements in this condition.

This redundancy argument predicted in advance that prosody should not influence dog behaviour in "sit" commands, but why did we find no effect for "stand" (i.e., we found an effect for "down" only)? An inspection of the data reveals one likely explanation: "down" commands in our dataset were almost universally delivered in isolation (96%; 108 of 112), while "stand" commands were most often preceded by the dog's name (66%; 45 of 68). This made prosody less informative for "stand" because the dog's name served as an earlier additional cue: put simply, dogs know to stand when they hear their name. In contrast, prosody is most informative in "down" commands, which is also where we found its significant influence on dog reactions. In the remaining commands, it is either fully non-informative ("sit"-up, "sit"-down) or less informative (in "stand").

The fact that dogs are phylogenetically distant from humans, but appear to share understanding of certain sound symbolic associations in vocal comprehension points to a distant but shared evolutionary origin. This origin may be one that endows innate cross-sensory associations per se, or simply an ability to extract environment statistics (since dogs typically share the human environment from birth). So it is interesting to consider exactly *why* pitch signals elevation to humans and dogs (which we see both in cross-modal correspondences[9,10,13], and the sound symbolism of certain words here). Often these associations are not random: for example, high-pitch is associated with small and elevated, likely because high pitch is emitted from smaller animals and from objects higher in space, in scene statistics (e.g[32]). Another contributing factor may be that low-pitched voices in humans and other mammals signals dominance[33] (which triggers submissive downward-type postures like dropped head, lowered ears, and tucked tail;[34,35]). Importantly, this is still sound symbolism: the dog is responding appropriately to the meaning of the word "down" (lying down) and doing so more quickly depending on how the word is pronounced. We also have evidence that our effect cannot be explained by dominance alone. In a dominance-only account, our findings could be nothing more than humans expressing dominance using low pitch sounds, and dogs assuming low (submissive) positions -- without any link to language. But this simplistic account does not fit the data. It would predict that dogs move down faster to all low pitch commands, but they did not (i.e., pitch did not affect their response to "sit" down commands, as expected from a sound symbolism account– see above). It would also predict, contrary to our results, that standing should be harder with high pitch "stand", since high pitch signals a small/safe speaker which typically elicits approach rather than upright stillness. Finally, it would predict that professional handers should use lower pitch than amateurs, because professionals are more dominant (given their greater need to control the dog's behaviour). But our data show the opposite: professionals used higher pitch than amateurs and were more willing to vary their pitch to signal elevation (e.g., they raised pitch for "sit" up). In other words professionals were not simply more dominant (as a pure dominance account would predict) but were instead better able to use pitch to convey linguistic meaning. In summary, dogs are responding to the linguistic environment in a way that aligns with sound symbolism (lowering faster to low pitch "down"), and this may potentially stem from two sources: the known scene statistics linking pitch and elevation [e.g., 27] and perhaps also the notion of dominance reinforcing this association between pitch and word-meaning.

Our study has several limitations, for example we could not control the competition arena. Furthermore, our study is correlational so we must take care when inferring causality; there may yet be other factors beyond pitch (see [36] which affected dogs' performance in our study (although we have already addressed a number of these in our discussion above). Future studies might also seek to understand why our behavioural effect was in duration-of-movement but not latency-to-move. One possibility is ceiling effects because our competition dogs were high-level performers. Future studies might find stronger effects in younger, less experienced dogs, or in puppies (who are known to show preferences for prosody in dog-directed speech[37,38]. We also note that future studies might expand on our analyses of mean f_0 to consider f_0 contours, which could themselves be implicated in prosodic sound symbolisms. Two stimuli with different contours or f_0 range could have similar f_0 mean, so a contour analysis may reveal further effects not apparent from mean alone.

In conclusion, dog owners employed sound symbolic prosody, speaking with higher mean pitch in "stand" compared to "down" (and also in "sit"-up vs. "sit"-down for the most advanced competition handlers). Our data relate to findings from pitch-elevation correspondences in animals [25], and sound symbolic prosody between humans[9], but show a crucial degree of mutual inter-species understanding for sound symbolic language in certain circumstances, never previously seen. Handlers not only employed referential prosody, but dogs appeared to understand this, reacting faster when "down" commands were uttered in lower pitch. Prosodic sound symbolism may therefore facilitate mutual understanding in some areas of cross-species communication, and prove a useful training strategy in competition.

Methods

Study 1: acoustic analysis of prosody in voice commands from amateurs

Materials and procedure

Thirty-four participants completed an online questionnaire and six spoken exercises. Our questionnaire elicited (i) the owner's dog-owning experience (in months), (ii) owner's training experience (none, basic obedience training only, training for sport/ competition, professional) (iii) dog's overall level of training (informal, formal, sport training) and (iv) dog's competence in the commands of "sit", "down" and "stand" (responses on a 4-point scale; see Design). Supplementary Information (SI) shows the full questionnaire (including additional items to be examined elsewhere).

Participants then recorded six spoken commands given to their dog, without hand signals, using a downloaded audio recorder. These commands are shown in Table 1 (i.e., two instances of "stand", "down", "sit"), and their order was randomised for each participant. Voice file recordings were spliced into individual commands (excluding any dog name or other vocal cue). We used the Voice Report function in PRAAT with a pitch range of 75–700 Hz to extract the mean f_0 for each voice command[39]. Approximately 20% of voice files were double-coded to ensure accuracy (Cronbach's alpha 0.99, 95% CIs [0.999, 1]. In < 3% of utterances, owners used different command words for the same instructions, as follows: (i) "stand" was sometimes "back"; both were collapsed because they contain the same vowel quality (/stand/= /bak/). One participant used "up", which has a different vowel quality but their data was retained since pitch fell within 2 SD from the mean of the remanding items. (ii) "down" was sometimes "lie" (/laɪ/ or /lʌɪ), and again these instances were collapsed with "down" since they start from approximately the same position in vowel space (cf., /daʊn/)[40].

Study 2: Self-report of prosody in voice commands from competitors

Materials and procedure

One hundred and seventy participants completed an online questionnaire[41]. We asked whether vocal utterances for "stand" and "down" were made with high (or rising) pitch versus low (or falling) pitch (henceforth high or low, respectively). Two further questions elicited from owners the number of dogs trained in Competition Obedience, and highest class in competition (with 8 response categories from *Pre-beginner* to *Crufts Championship*). Remaining questions asked how participants met Kennel Club regulations for competition, i.e., whether they used the dog's name before the command (this is optional), and whether they use (a) body signals, versus (b) vocal commands (e.g., "stand") – since Kennel Club regulations specify they cannot use both. See SI for full questionnaire (including additional items to be examined elsewhere, e.g., regarding loudness).

Study 3

Study 3a: acoustic analysis of prosody in voice commands in competition

We analysed publicly available videos of the Crufts Distant Control competition between 2012 and 2018 inclusive. Seventy-one target videos met our requirements that the handler was visible, gave spoken commands (not hand signals) and did not use dog's name in the command. From these, we analysed mean f_0 for each of the six commands (see Table 1) using PRAAT's Voice Report function (100–850 Hz pitch range[39]. In supplementary analyses we added an additional 107 commands, which we extracted from instances where the owner had also used the dog's name before the command (analysing the mean pitch (f_0) from the command alone). Approximately 15% of voice files were double-coded to ensure accuracy (Cronbach's alpha 0.99, 95% CIs [0.999, 0.999]).

Study 3b: analysis of dogs' performance as a function of prosody in voice commands

We selected 68 videos from Study 3a, which were where both handler *and dog* were visible (since we would be coding dogs' behaviour). Once again, we avoided using the videos in which handlers used hand signals to give commands. Our choice of participant group (not using hand signals) and our choice of competition (distant control) therefore both ensured again that no extraneous factors other than the voice command would dictate the dogs' behaviour. Videos were muted and cropped to show only the dog, then coded using BORIS (Behaviour Observation Research Interactive Software v.7.10.5[42] for two timed behaviours: (a) *latency-to-move*: time in seconds between the onset of the voice command (signalled by a chime we inserted into our videos to blind-code them) and the onset of the dog's movement (b) *duration-of-movement*: time between the moment the dog moved, and the moment it settled into its final position (e.g., sitting). Approximately 30% of voice files were double-coded to ensure accuracy, giving Cronbach's alpha 0.99 for latency (95% CIs [0.87, 0.99]) and 0.96 for duration-of-movement (95% CIs [0.85, 0.99]).

Ethics

The University of Sussex Ethical Review Committee approved Studies 1 and 2 (ER/AK682/3). Study 3 used publicly available video materials. Informed consent was obtained from all participating dog owners. In the design of the experiments involving dogs we followed the ASAB guidelines for ethical treatment of non-human animals in research[43]. Research involving human research participants was performed in accordance with the BPS Code of Human Research Ethics[44] and the Declaration of Helsinki[45].

Data availability

The data underlying this publication is openly available on OSF and can be accessed here: [https://osf.io/3xwn4/](https:/osf.io/3xwn4).

Received: 28 January 2025; Accepted: 3 October 2025
Published online: 11 November 2025

References

1. Nielsen, A. K. S. & Dingemanse, M. Iconicity in word learning and beyond: A critical review. *Lang. Speech.* **64** (1), 52–72. https://doi.org/10.1177/0023830920914339 (2021).
2. Bankieris, K. & Simner, J. What is the link between synaesthesia and sound symbolism? *Cognition* **136** https://doi.org/10.1016/j.cognition.2014.11.013 (2015).
3. Birch, D. & Erickson, M. Phonetic symbolism with respect to three dimensions from the semantic differential. *J. Gen. Psychol.* **58** (2), 291–297. https://doi.org/10.1080/00221309.1958.9920401 (1958).
4. Sapir, E. The study in phonetic symbolism. *J. Exp. Psychol.* **12** (3), 225–239. https://doi.org/10.1515/9783110198867 (1929).
5. Köhler, W. Gestalt psychology. In *Gestalt psychology.* Liveright. (1929).
6. Ramachandran, V. S., Marcus, Z. & Chunharas, C. Chapter 1 - Bouba-Kiki: cross-domain resonance and the origins of synesthesia, metaphor, and words in the human Mind. In Multisensory Perception. Elsevier Inc. https://doi.org/10.1016/B978-0-12-812492-5.00001-2 (2000).
7. Margiotoudi, K. et al. Bo-NO-bouba-kiki: Picture-word mapping but no spontaneous sound symbolic speech-shape mapping in a Language trained Bonobo. *Proc. Royal Soc. B: Biol. Sci.* **289** (1968). https://doi.org/10.1098/rspb.2021.1717 (2022).

8. Nygaard, L. C., Herold, D. S. & Namy, L. L. The semantics of prosody: acoustic and perceptual evidence of prosodic correlates to word meaning. *Cogn. Sci.* **33** (1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x (2009).
9. Shintel, H., Nusbaum, H. C. & Okrent, A. Analog acoustic expression in speech communication. *J. Mem. Lang.* **55** (2), 167–177. https://doi.org/10.1016/j.jml.2006.03.002 (2006).
10. Spence, C. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **73** (4), 971–995. https://doi.org/10.3758/s13414-010-0073-7 (2011).
11. Pratt, C. C. The Spatial character of high and low tones. *J. Exp. Psychol.* **13** (3), 278–285. https://doi.org/10.1037/h0072651 (1930).
12. Ludwig, V. U., Adachi, I. & Matsuzawa, T. Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661–20665. (2011). https://doi.org/10.1073/pnas.1112605108
13. Korzeniowska, A. T., Root-Gutteridge, H., Simner, J. & Reby, D. Audio–visual crossmodal correspondences in domestic dogs (Canis familiaris). *Biol. Lett.* **15** (11). https://doi.org/10.1098/rsbl.2019.0564 (2019).
14. Loconsole, M., Pasculli, M. S. & Regolin, L. Space-luminance crossmodal correspondences in domestic chicks. *Vision. Res.* **188** (January), 26–31. https://doi.org/10.1016/j.visres.2021.07.001 (2021).
15. Loconsole, M., Stancher, G. & Versace, E. Crossmodal association between visual and acoustic cues in a tortoise (Testudo hermanni). *Biol. Lett.* **19** (20230265). https://doi.org/10.1098/rsbl.2023.0265 (2023).
16. The Kennel Club. *Obedience Regulations* (2022).
17. Ewan, W. G. Explaining the intrinsic pitch of vowels. *J. Acoust. Soc. Am.* **58** (S1), S40–S40. https://doi.org/10.1121/1.2002115 (1975).
18. R Core Team. *R: A language and environment for statistical computing* (4.1.2). (2022). https://www.r-project.org/
19. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear Mixed-Effects models using lme4. *J. Stat. Softw.* **67** (1), 1–48 (2015).
20. Kassambara, A. *ggpubr: ggplot2 Based Publication Ready Plots. R package version 0.4.0.* (0.4.0). (2020).
21. Halekoh, U., Hojsgaard, S. & Yan, J. The R package geepack for generalized estimating equations. *J. Stat. Softw.* **23**, 859–880 (2006).
22. Anrep, G. V. Pitch discrimination in the dog. *J. Physiol.* **53** (6), 367–385 (1920).
23. Barber, A. L. A. et al. A comparison of hearing and auditory functioning between dogs and humans. *Comp. Cognit. Behav. Rev.* **15**, 45–94. https://doi.org/10.3819/ccbr.2020.150007 (2020).
24. Heffner, H. E. & Heffner, R. S. High-Frequency hearing. *Senses: Compr. Ref.* **3** (January 2019), 55–60. https://doi.org/10.1016/B978-012370880-9.00004-9 (2008).
25. Korzeniowska, A. T., Simner, J., Root-Gutteridge, H. & Reby, D. High-pitch sounds small for domestic dogs: abstract crossmodal correspondences between auditory pitch and visual size. *Royal Soc. Open. Sci.* **9** (2). https://doi.org/10.1098/rsos.211647 (2022).
26. McConnell, P. B. Lessons from animal trainers: the effect of acoustic structure on an animal's response. In *Perspectives in Ethology* (pp. 165–187). (1991).
27. Blake, H. *Talking with Horses: A Study of Communication Between Man and Horse*. Souvenir. (2001). https://books.google.co.uk/books?id=n-AJPwAACAAJ
28. Holmes, J. *The Farmer's Dog*. Popular Dogs. (1984). https://books.google.co.uk/books?id=xPeiwfYtyXcC
29. Krebs, J. R. & Dawkins, R. Animal signals: Mind-Reading and manipulation. In (eds Krebs, J. R. & Davies, N. B.) Behavioural ecology. An Evolutionary Approach (380–401). Blackwell Scientific. (1984).
30. McConnell, P. B. & Baylis, J. R. Interspecific communication in cooperative herding: acoustic and visual signals from human shepherds and herding dogs. *Z. Fur Tierpsychologie*. **67**, 302–328 (1985).
31. Woodhouse, B. *Dog Training my Way* (Berkley Books, 1984). https://books.google.co.uk/books?id=jOTi1tuIqjgC
32. Parise, C. V., Knorre, K. & Ernst, M. O. Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108. (2014). https://doi.org/10.1073/pnas.1322705111
33. Puts, D. A., Hodges, C. R., Cárdenas, R. A. & Gaulin, S. J. C. Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evol. Hum. Behav.* **28** (5), 340–344. https://doi.org/10.1016/j.evolhumbehav.2007.05.002 (2007).
34. Morton, E. S. On the occurrence and significance of Motivation-Structural rules in some bird and mammal sounds. *Am. Nat.* **111** (981), 855–869. https://doi.org/10.1086/283219 (1977).
35. Reby, D. & McComb, K. Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Anim. Behav.* **65** (3), 519–530. https://doi.org/10.1006/anbe.2003.2078 (2003).
36. Mills, D. S. What's in a word? A review of the attributes of a command affecting the performance of pet dogs. *Anthrozoos* **18** (3), 208–221. https://doi.org/10.2752/089279305785594108 (2005).
37. Ben-Aderet, T., Gallego-Abenza, M., Reby, D. & Mathevon, N. Dog-directed speech: why do we use it and do dogs pay attention to it? *Proc. Royal Soc. B: Biol. Sci.* **284** (1846). https://doi.org/10.1098/rspb.2016.2429 (2017).
38. Benjamin, A. & Slocombe, K. Who's a good boy?!' dogs prefer naturalistic dog-directed speech. *Anim. Cogn.* **21** (3), 353–364. https://doi.org/10.1007/s10071-018-1172-4 (2018).
39. Boersma, P. & Weenink, D. *Praat: doing phonetics by computer* (6.3.05). (2023).
40. Ladefoged, P. *A Course in Phonetics* (Harcourt College, 2001).
41. Qualtrics. *Qualtrics, Provo* (UT, 2021).
42. Friard, O. & Gamba, M. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* **7** (11), 1325–1330 (2016).
43. ASAB Ethical Committee & ABS Animal Care Committee. Guidelines for the ethical treatment of nonhuman animals in behavioural research and teaching. *Anim. Behav.* **195** (I-XI). https://doi.org/10.1016/j.anbehav.2022.09.006 (2023).
44. Oates, J. et al. *BPS Code of Human Research Ethics*. (2021). https://doi.org/10.53841/bpsrep.2021.inf180
45. World Medical Association. World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *J. Am. Med. Association*. **310** (20). https://doi.org/10.1001/jama.2013.281053 (2013).

Acknowledgements
We are very grateful to all the dog owners and dogs who took part in our studies.

Author contributions
ATK, DR and JS conceived the project. ATK ran the study, conducted the analyses, and drafted Methods/Results. JS and DR supervised the project. JS wrote the manuscript. DR/ATK/ HRG edited the manuscript. HRG assisted with acoustic analyses and conducted the second coding.

Declarations

Competing interests
The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-22923-1.

Correspondence and requests for materials should be addressed to J.S.