



Is it all about the pitch? Acoustic determinants of dog-directed speech preference in domestic dogs, *Canis familiaris*

Anna Gergely^{a,*}, Katinka Tóth^a, Tamás Faragó^b, József Topál^a

^a Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary

^b Department of Ethology, Eötvös Loránd University, Budapest, Hungary

ARTICLE INFO

Article history:

Received 21 October 2020

Initial acceptance 30 November 2020

Final acceptance 23 February 2021

Available online 11 May 2021

MS number 20-00778R

Keywords:

dog-directed speech
hemispheric lateralization
infant-directed speech
prosody processing

Dogs, similarly to infants, have been shown to be sensitive to human speech especially when it is directed to them. However, what essential acoustic, paralinguistic and lexical features of dog-directed speech are responsible for this preference in dogs is largely unknown. In the present study, generalized dog (DDS)-, infant (IDS)- and adult (ADS)-directed speech stimuli were created by using prerecorded sentences of multiple female speakers and these composite (averaged) stimuli were then manipulated to control for linguistic content as well as to equalize their mean fundamental frequency (F0) value. All three possible pairwise combinations of these acoustic stimuli were then presented to adult dogs in a two-way choice task where two identical target objects were used to indicate the sound sources. We found a significant preference towards the target object associated with DDS in the DDS versus ADS condition and suggest that, for dogs, mean F0 difference is not essential for DDS–ADS discrimination. However, we did not find evidence of selection bias when IDS was simultaneously presented either with DDS or ADS. Interestingly, our results also showed that dogs were more willing to approach the ‘more prosodic’ location (i.e. DDS or IDS versus ADS) when the prosodically more prominent sound stimulus was presented on their left side which suggests right-hemispheric specialization for neural processing of prosodic sounds in this domestic species. We also found that dogs made their choice faster when the ‘more prosodic’ stimulus was given first which suggests that they can perceive the difference not only between DDS and ADS, but also between IDS and ADS and between IDS and DDS. In conclusion, the composite DDS, IDS and ADS stimuli in the present study proved to be an effective technique in exploring the acoustic determinants of dog-directed speech preference in dogs.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Acoustic and linguistic features of the language spoken by adults usually depends a lot on the addressee and his/her language comprehension skills. People, for example, tend to use a specific register when they speak to a preverbal infant (infant-directed speech, IDS). This type of speech is characterized by exaggerated contouring of fundamental frequency (F0, perceived as pitch), higher absolute F0, wider F0 range, altered duration of vocalizations and pauses, stricter tempo, greater repetition, vowel hyper-articulation and simplified syntax compared to adult-directed speech (ADS; e.g. Burnham, Kitamura, & Vollmer-Conna, 2002; Fernald, 1989; Stern, Spieker, & MacKain, 1982). The prosodic properties of IDS have two important functions: (1) the acoustic features (an increase in the F0, wide F0 range, exaggerated F0

contour, etc.) serve to capture and maintain infants' attention whereas (2) the paralinguistic characteristics (e.g. vowel hyper-articulation, repetition, slower tempo) facilitate language learning (e.g. Cooper, Abraham, Berman, & Staska, 1997; Song, Demuth, & Morgan, 2010). Importantly, mothers spontaneously adjust various aspects of their IDS as a function of their infants' need and language ability. For example, they use less exaggerated acoustic prosody towards children with more advanced language comprehension skills (Liu Tsao, 2009).

It has also been demonstrated that infants show clear preference at both behavioural and neural levels towards the speech that is directed to them (e.g. Cooper & Aslin, 1990; Fernald, 1985; Naoi, Minagawa-Kawai, Kobayashi, Takeuchi, & Nakamura, 2012; Sulpizio et al., 2018). Acoustic, paralinguistic and linguistic determinants of IDS that are essential for eliciting infants' preference have been studied in detail (e.g. Fernald & Kuhl, 1987; Nencheva, Piazza, & Lew-Williams, 2020). In their seminal study, Fernald

* Corresponding author.

E-mail address: gergely.anna@ttk.hu (A. Gergely).

and Kuhl (1987) used manipulated (i.e. sine wave) speech signals and found that 4-month-old infants show a preference for IDS over ADS only if the signal is characterized by a specific F0 pattern (i.e. mean F0, F0 range and contour). Linguistic content, amplitude and temporal pattern, however, play only a minor role in capturing infants' attention. In line with these results, Nenchewa, Piazza, and Lew-Williams (2020) provided evidence that children's attention dynamics (measured in terms of the changes in pupil size) is aligned with the F0 contour of IDS. They also found that stimuli with specific IDS contour (i.e. 'fall' and 'hill' patterns) can capture and maintain infants' attention more efficiently than stimuli with other types of IDS contours (i.e. 'valley' and 'rise' patterns) or stimuli with typical ADS contour.

Behavioural preference towards addressee-specific speech (i.e. dog-directed speech, DDS) has also been shown in dogs (Benjamin & Slocombe, 2018; Jeannin, Gilbert, Amy, & Leboucher, 2017), but the role of the acoustic and paralinguistic features behind this preference is still largely unknown. Jeannin et al. (2017), for example, reported that elevated mean F0 is an essential acoustic determinant of dogs' preference for DDS over ADS and that adult dogs' attention showed positive correlation with F0 mean. However, other acoustic parameters of the speech registers, like F0 range, intonation contour (i.e. difference between the ending and starting F0) and harmonicity seemed to have no effect on adult dogs' and puppies' attention (Jeannin et al., 2017). In contrast, another study found a correlational effect of the F0 mean and dogs' attention only in puppies but not in adult dogs and concluded that adult dogs showed reduced willingness to respond to human verbal play signals (Ben-Aderet, Gallego-Abenza, Reby, & Mathevon, 2017). These inconsistencies may stem from methodological differences between the two aforementioned studies as Jeannin et al. (2017) recoded the acoustic stimuli while speakers were talking to live partners while the other study used sound recordings from speakers that were talking to pictures of their partners (Ben-Aderet et al., 2017). We may also assume that the lexical content of a given speech stimulus can also affect dogs' responses. There is only one study examining the effects of congruent/incongruent lexical content of DDS/ADS on dogs' preference (Benjamin & Slocombe, 2018) and this suggests a combined role for congruent dog-directed prosody and lexical content in dogs' preferential attention to DDS. Therefore, it is also possible that resolution of the aforementioned, seemingly contradictory results lies in the systematic differences between lexical and contextual information used in stimulus playbacks (i.e. multiple fixed playful sentences (Ben-Aderet et al., 2017) versus one fixed sentence about going for a walk (Jeannin et al., 2017)). Beyond prosodic and linguistic features of DDS, the speakers' identity can also be important for dogs when hearing dog-directed acoustic stimuli (e.g. Benjamin & Slocombe, 2018).

There is also emerging evidence that DDS differs not only from ADS, but also from IDS. Natural DDS (i.e. directed to the speaker's own family dog) is characterized by higher F0 than IDS (Gergely, Faragó, Galambos, & Topál, 2017). Furthermore, certain paralinguistic features of IDS (vowel hyperarticulation) seem to be missing from DDS (Gergely et al., 2017; Jeannin et al., 2017; Xu, Burnham, Kitamura, & Vollmer-Conna, 2013). This can be interpreted as indicating that towards nonverbal listeners, such as dogs, we aim to use an exaggerated attention-getting but not language-tutoring speech style. Despite these differences between DDS and IDS, it has been shown that dogs respond similarly to IDS and DDS (Jeannin et al., 2017). Interestingly, dogs' responses are also similar towards IDS and ADS at a behavioural level, which is surprising considering the striking acoustic differences between the two speech styles (Jeannin et al., 2017). These authors reported that the IDS stimuli used in their study had greater intensity modulation than the DDS and ADS stimuli (Jeannin et al., 2017), but did not

discuss how this can result in similar responses to ADS–IDS and DDS–IDS in dogs.

The aim of the present study was to investigate whether dogs' preference towards DDS can be elicited in the absence of its high overall pitch (mean F0) and lexical content. To do so, we created composite DDS, IDS and ADS stimuli that have similar overall pitch (mean F0) without manipulating any other prosodic features directly. To eliminate any possible effect of lexical content, we generated sine waves based on the F0 contour of the sentences which eliminated the formant structure whereas the prosodic features remained unchanged (similarly to Fernald & Kuhl, 1987; Ratcliffe & Reby, 2014). Controlling for the speaker's identity, we did not use one particular sentence from one speaker, but generated a composite DDS/IDS/ADS stimulus by using the same sentence of multiple speakers (for details see Methods). We used only female voices to make our study comparable to previous studies that focused on DDS preference in dogs (Ben-Aderet et al., 2017; Benjamin & Slocombe, 2018; Jeannin et al., 2017). Dogs in the present study were presented with these general DDS–ADS, DDS–IDS and IDS–ADS stimulus pairs in a two-way choice task in which two identical target objects were presented. We hypothesized that, despite elimination of mean F0 differences, the generated DDS would still contain a sufficient amount of prosodic information to make this representative averaged DDS distinguishable from ADS; therefore, we predicted that dogs would show a preference towards DDS over ADS. At the same time, we can assume that our method of creating composite stimuli and the elimination of F0 mean difference makes it even more challenging for the dogs to distinguish between DDS–IDS and IDS–ADS (Jeannin et al., 2017); therefore, they were expected to show a similar response when faced with DDS versus IDS and IDS versus ADS pairs.

METHODS

Ethical Note

This research was approved by the National Animal Experimentation Ethics Committee (Ref. No. PEI/001/1057–6/2015). Research was done in accordance with the Hungarian regulations on animal experimentation and the ASAB/ABS Guidelines for the use of animals in research.

Subjects

We recruited 65 adult family dogs through the database of the Family Dog Project at Eötvös Loránd University, Budapest, Hungary and by using an online call in a closed Hungarian Facebook group called 'Canine Ethology' operated by employees of the Family Dog Project at Eötvös Loránd University and Hungarian Academy of Sciences. Dogs had to be older than 1 year and to be motivated to play with tennis balls. Five dogs were excluded from the final analysis because they did not approach the tennis balls within 30 s after release in at least one of the test trials (see Procedure). The remaining 60 dogs (mean age 5.1 ± 2.8 years, 31 females, 29 males) were included in the statistical analysis (20 dogs in each condition, see below). Each dog participated in only one condition. In the DDS versus ADS condition there were one akita, one bichon havanese, one boxer, two cairn terriers, one corgi, one German shepherd, one groenendael, one Hungarian vizsla, one münsterländer, one puli, one Shetland sheepdog, one shiba inu, two schipperkes, one whippet and four mongrels (mean age \pm SD: 5.2 ± 3.1 years, nine females, 11 males). In the DDS versus IDS condition there were one Australian kelpie, two beaucherons, one Belgian malinois, three golden retrievers, three labrador retrievers, two mudis, one puli

and seven mongrels (mean age \pm SD: 4.9 ± 3 years, 10 females, 10 males). In the IDS versus ADS condition there were one cavalier King Charles spaniel, one German shepherd, two golden retrievers, one groenendael, two Hungarian vizslas, one Parson Russel terrier, two Siberian huskies and 10 mongrels (mean age \pm SD: 5.2 ± 2.5 years, 12 females, eight males).

Stimuli Preparation

First, we chose infant- adult- and dog-directed versions of the same Hungarian sentence ('No nézd csak, milyen szép idő van odakint!', in English: 'Just look outside, what nice weather!') from six female speakers when addressing their 0–8-month-old infants (IDS), their own adult family dogs (DDS) and an adult female experimenter (ADS). These recordings were originally collected for other research purposes (for details see Gergely et al., 2017). This exact sentence was recorded twice from all six speakers in all three conditions. Therefore, this procedure resulted in 2×6 DDS, 2×6 IDS and 2×6 ADS recordings, 36 sentences in total.

These original recordings were then processed with PRAAT software (version 6.0.05, <http://www.praat.org>) to create acoustic stimuli for the present study. First, all speakers' sentences were annotated; then we extracted voiced parts (calls: 'no né a ilyen é i dővano da in') from each recording. F0 contours were then extracted from each section. Next, lengths of each matched section were averaged across speakers. To eliminate the effect of F0 mean difference in speech registers, the F0 mean of the DDS, IDS and ADS stimuli was shifted to 220 Hz (mean F0 of female voice, Pisanski et al., 2016; Titze, 2000; see Fig. 1). Then matching sections' contours within DDS, IDS and ADS were averaged across speakers, and sinus sounds were generated from the F0 contour to eliminate lexical content as well. As a final step, section onsets were matched with those of one reference speaker (age 27) to mimic normal speech dynamics in DDS, IDS and ADS separately and then these stimuli were normalized to the same average sound level (-27 dB RMS) to prevent dogs from listening for an overall level difference. Following this procedure, we generated three composite stimuli (1-1-1 DDS, ADS and IDS, respectively) that possessed similar mean F0 and intensity but contained averaged and not directly manipulated

F0 contour and range, call length, rhythm and speech dynamics, etc. without linguistic content (see Supplementary audio samples and Table 1). These DDS, IDS and ADS samples were used in pairs in the present playback experiments. We created only one DDS, one ADS and one IDS stimulus that were averaged, sine waved and F0 equalized; therefore, every dog heard the same DDS, ADS and IDS stimulus during the test phase.

This stimulus manipulation procedure (multistep modification of groups of natural DD, ID and AD sound stimuli) resulted in artificial stimuli that may be considered representative of their respective broader categories (infant- dog- and adult-directed speech). As a result of our stimulus generalization process, acoustic parameters of the generalized sounds deviated less from the mean, and thus contained more homogeneous sound segments, than the original sentences (see Table 1). Moreover, variance caused by individual speech style (individual tone, pitch, rhythm, etc.) was greatly reduced in the composite (averaged) stimuli, while typical dog-, infant- and adult-directed features of the speech prosody remained intact (for details see Table 1).

Experimental Arrangement

The experiment was conducted in a laboratory room (5×2.5 m) with tape on the floor marking standardized locations of the experiment (Fig. 2). Video cameras were mounted on each wall, with output recorded on computer. Two identical loudspeakers (Logitech X-230 2.1), used for audio stimulus playbacks, were placed as far as possible from each other (160 cm) to make it easy for the dog to tell whether the sound came from the left or the right loudspeaker (see Fig. 2). We used two identical yellow tennis balls as target objects.

For the three experimental conditions we created three sound stimuli pairs from the generated sine-waved samples (see above): DDS versus ADS (DDS–ADS); DDS versus IDS (DDS–IDS); IDS versus ADS (IDS–ADS). During the test phase one stimulus from the pair came from one loudspeaker (e.g. left) while the other came from the other loudspeaker (e.g. right). Dogs are known to habituate easily and quickly lose interest in the stimuli in such playback experiments (e.g. Jeannin et al., 2017); thus, we gave each dog only

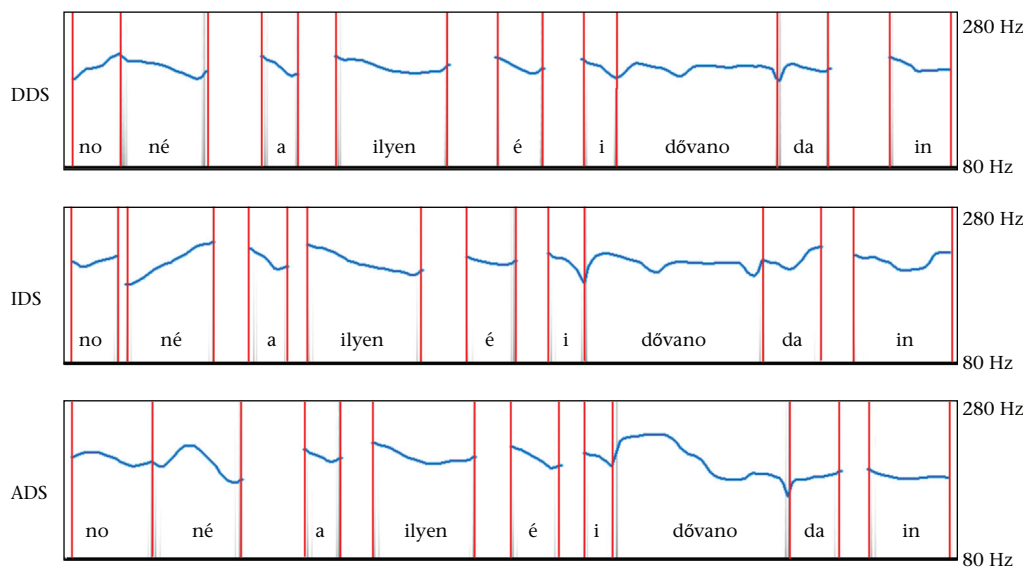


Figure 1. F0 contours of the average dog-directed (DDS), infant-directed (IDS) and adult-directed (ADS) stimuli. All three sound samples were shifted to 220 Hz to eliminate the F0 difference between them. Red lines indicate the original call segments that were replaced by sinusoid sounds to eliminate lexical content of the stimuli. The original sentence in Hungarian was 'No, nézd csak, milyen szép idő van odakint!'.

Table 1
Acoustic parameters of the original sentences and generated sound stimuli used in the present experiment

	Original sentences			Generalized stimuli			Cause of difference
	DDS	IDS	ADS	DDS	IDS	ADS	
F0 mean (Hz)	215.8	229.8	187.8	220.0	220.0	220.0	F0 mean shifting
	<i>46.2</i>	<i>48.5</i>	<i>42.4</i>	<i>0.6</i>	<i>0.7</i>	<i>7.3</i>	
Intensity mean (dB)	65.1	64.3	59.9	81.6	81.6	81.6	Normalization
	<i>7.0</i>	<i>7.7</i>	<i>5.7</i>	<i>0.2</i>	<i>0.2</i>	<i>0.3</i>	
F0 maximum (Hz)	244.5	267.4	213.3	233.4	234.5	230.0	Sound merging (averaging) & sinus sound preparation
	<i>53.4</i>	<i>58.3</i>	<i>55.5</i>	<i>3.0</i>	<i>7.2</i>	<i>11.9</i>	
F0 minimum (Hz)	193.9	204.1	166.0	209.6	206.6	203.0	
	<i>41.7</i>	<i>47.6</i>	<i>42.6</i>	<i>7.3</i>	<i>11.8</i>	<i>13.9</i>	
F0 range mean (Hz)	50.6	63.3	47.3	23.8	27.9	27.0	
	<i>33.9</i>	<i>46.1</i>	<i>46.2</i>	<i>8.1</i>	<i>17.2</i>	<i>20.9</i>	
Call length mean (s)	0.20	0.25	0.17	0.20	0.24	0.21	
	<i>0.2</i>	<i>0.2</i>	<i>0.1</i>	<i>0.2</i>	<i>0.2</i>	<i>0.1</i>	
HNR mean (dB)	13.9	15.5	12.5	37.2	35.0	35.5	
	<i>4.5</i>	<i>4.4</i>	<i>4.8</i>	<i>5.3</i>	<i>4.3</i>	<i>6.6</i>	
HNR deviation (dB)	5.0	4.7	4.6	12.0	11.9	11.6	
	<i>1.4</i>	<i>1.9</i>	<i>2.1</i>	<i>1.7</i>	<i>2.3</i>	<i>2.2</i>	
Jitter mean (μ s)	0.011	0.009	0.011	0.004	0.004	0.004	
	<i>0.01</i>	<i>0.01</i>	<i>0.01</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	

F0 = fundamental frequency, HNR = harmonic to noise ratio. The mean of individual call segments is given with the SEM in italics. F0 shift and normalization were directly manipulated. All other feature differences were caused by merging (i.e. averaging) of the original sentences and lexical content elimination (i.e. sinusoid sound generation).

two test trials to examine dogs' spontaneous preference and to control for stimulus playback order. Sides and order of the sound playbacks were counterbalanced across subjects within and between conditions.

Procedure

Pretest phase

The owner and the dog entered the room with the experimenter (E). Then the dog was allowed to sniff and explore the room for 1 min. During this period E informed the owner about the procedure. Then E initiated ball play with the dog in the middle of the room by throwing each ball once and encouraging the dog to retrieve it. If the dog did not touch both balls, E threw them once again (the dogs had to touch both balls at least once).

Test phase

The owner sat down at a predetermined location and held the dog in front of him/herself (see Fig. 2). E held a tennis ball in each hand. She showed the balls to the dog, then stepped backwards and placed them on the ground in front of each loudspeaker. Since 160 cm was too wide for simultaneous placement, E put the balls down by the loudspeakers one after the other, then squatted halfway between the two balls and swung her arms while reaching towards each ball and touched both gently with the tip of her fingers without grabbing or lifting them. She did this two to four times until the dog looked at each ball at least once (the side of the last touched ball was randomized between and within subjects, i.e. between the two test trials). Then E walked back to the dog showing her empty hands and went into the adjoining computer room to replay the sound stimuli of a pair in succession with an interstimulus interval of 2 s. That is, one auditory stimulus (e.g. DDS) was played through one loudspeaker (e.g. left), and after 2 s of silence, the other stimulus (e.g. ADS) was played by the other loudspeaker (right). The owner then released the dog and encouraged it (saying e.g. 'You can go!', 'Let's go!'). If the dog chose one of the tennis balls (i.e. approached a tennis ball to within 30 cm) the owner praised the dog and it was allowed to play with the ball for a few seconds. Meanwhile E entered the room and collected both tennis balls then initiated play with the dog in the middle of the room by throwing each ball once again. The whole

procedure was repeated but this time we reversed the order of the stimulus presentation, while the side of the stimuli remained the same (e.g. if the stimulus presentation during the first trial was right-DDS and then left-ADS, dogs were presented with left-ADS and then right-DDS during the second trial).

Data Analysis

E coded the dog's choice during the experiment (i.e. she noted which tennis ball was chosen by the dog in each trial). The dog's behaviour was also analysed later with 0.2 s time resolution coding of all experimental recordings (with Solomon Coder, beta 16.06.26, <http://solomoncoder.com/>). Owing to technical failure, video recordings of seven dogs were damaged (two from DDS–ADS, three from DDS–IDS and two from IDS–ADS conditions); therefore, only the live-coded choice behaviour of these subjects was used in the analysis. The reliability of live coding of choice behaviour showed perfect agreement with video-based coding (Cohen's kappa coefficient: 1). To assess interobserver reliability, a second observer scored a randomly selected sample of 20% of recordings. Cohen's kappa coefficients (for categorical variables) and intraclass correlation coefficients (ICC, for continuous variables) are given below for each variable. The following behaviours were coded.

(1) Choice: a dog's choice behaviour was scored as 1 if it chose the tennis ball placed next to the 'more prosodic' sound source (i.e. the near-DDS tennis ball and the near-IDS tennis ball when IDS was contrasted with ADS) and 0 if it approached the tennis ball next to the 'less-prosodic' sound source (Cohen's kappa coefficient: 1).

(2) Latency of choice (s) was defined as the time elapsed between the moment when the owner released the dog and the moment when the dog approached a tennis ball within 30 cm with its nose (ICC: 0.88).

(3–4) Relative duration of looking towards the location of the 'more prosodic' sound source (%) was defined as the percentage of time spent looking towards the tennis ball next to the 'more prosodic' sound source (i.e. towards DDS versus IDS or ADS, and towards IDS versus ADS). This behaviour was coded separately during stimulus playback (i.e. during the first and second sound stimuli presentations) and during the choice phase, i.e. from the time of release until the dog approached one of the tennis balls within 30 cm (ICC: 0.92).

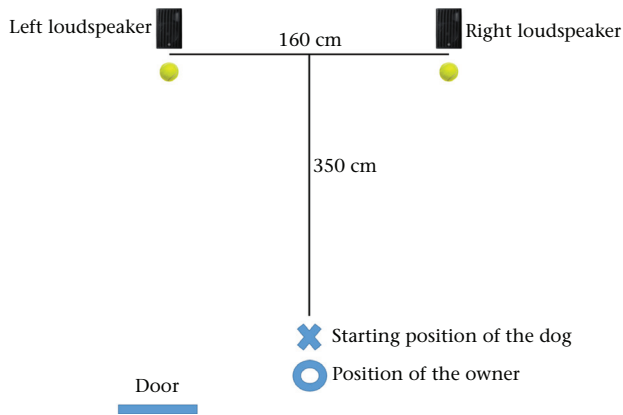


Figure 2. Experimental set-up. A yellow tennis ball was placed in front of each loudspeaker equidistant from the dog.

(5–6) Relative duration of looking towards the location of the ‘less prosodic’ sound source (%) was defined as the percentage of time spent looking towards the tennis ball next to the ‘less prosodic’ sound source (i.e. towards IDS or ADS versus DDS, and towards ADS versus IDS). This behaviour was also coded separately during stimulus playback (i.e. during the first and second sound stimuli presentations) and during the choice phase, i.e. from the time of release until the dog approached one of the tennis balls within 30 cm (ICC: 0.93).

First, dogs’ choice behaviour was analysed with one-sample binomial tests to examine whether they preferred to choose the target object next to the ‘more prosodic’ sound source in each test trial and experimental condition separately (chance level: 0.5). Next, we applied a binomial generalized linear mixed model (GLMM) for the choice variable using SPSS software version 22 (SPSS Inc., Chicago, IL, U.S.A.). Dogs’ looking behaviour towards the locations of ‘more prosodic and less prosodic’ sound sources during the stimulus playbacks and during choice was also analysed with paired-sample *t* tests separately in each trial. Third, a mixed-effects Cox regression model (MECRM, *coxme* package) was used for latency of choice analyses with R software (The R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org>). For MECRM, the hazard ratio ($\exp[\beta]$) between levels of a given fixed effect with 95% confidence interval is given. Subjects’ identities were included as a random grouping factor in all models to control for repeated measurements.

In GLMM and MECRM, the fixed explanatory variables were Condition (DDS–ADS, DDS–IDS, IDS–ADS), Trial (first, second), Stimulus order (more prosodic first, more prosodic second), Stimulus location (more prosodic on the left, more prosodic on the right) and all possible two-way interactions. In the MECRM, dogs’ choices were included in the model as a fixed explanatory variable (and all two-way interactions with choice and the explanatory variables) to investigate whether dogs chose faster when choosing the ball associated with the ‘more prosodic’ stimulus. The binomial model was not overdispersed. All tests were two tailed and the α value was set at 0.05. A sequential Bonferroni correction was applied in all post hoc comparisons. Nonsignificant interactions and main effects were removed from the model in a stepwise manner (backward elimination technique).

RESULTS

Choice Behaviour

Dogs preferred to choose the near-DDS tennis ball over the near-ADS one in the first trial (one-sample binomial test: DDS–ADS

condition: $N = 20$, $P = 0.04$). While they chose between the two options at chance level in the second trial ($P = 0.5$; Fig. 3). They also did not show a selection bias in DDS–IDS and IDS–ADS conditions in either trial (one-sample binomial tests: $N = 20$, 20, all $P \geq 0.5$; Fig. 3).

A binomial GLMM revealed that dogs’ choice behaviour was not influenced by any of the interactions, and these were removed from the model (Condition*Trial, Condition*Stimulus order, Condition*Stimulus location, Trial*Stimulus order, Trial*Stimulus location, Stimulus order*Stimulus location: all $P > 0.1$). Trial, Condition and Stimulus order also did not affect dogs’ choices as main effects. Therefore, these were also removed from the model (all $P > 0.1$). Stimulus location did have an effect on choice as dogs preferred to choose a tennis ball next to the ‘more prosodic’ sound source in all conditions but only when it was placed on the left side ($F_{1,118} = 5.1$, $P = 0.026$; Fig. 4).

The latency of choice MECRM revealed no significant interaction between any of the fixed effects ($N = 53$, all $P > 0.1$). As a main effect, Condition, Trial, Stimulus location and Choice had no influence on the latency of choice (all $P > 0.1$). At the same time, latency to choose was affected by Stimulus order (MECRM: $\chi^2_2 = 5.31$, $P = 0.021$). Dogs took less time to choose a tennis ball in general when hearing the ‘more prosodic’ stimulus first as opposed to hearing it second ($\exp(\beta) = 0.589$ [0.373; 0.929], $z = -2.27$, $P = 0.023$; Fig. 5).

Looking Behaviour

Dogs looked at the ‘more’ and ‘less’ prosodic sides equally long during stimulus playbacks and during choice in both trials in all three conditions (paired-sample *t* tests: DDS–ADS $N = 18$, DDS–IDS $N = 17$, IDS–ADS $N = 18$, all $P > 0.05$).

DISCUSSION

In the present experiment we found evidence that adult dogs show spontaneous preference towards the target object (tennis ball) associated with DDS over an identical tennis ball associated with ADS. This was so despite the lexical content and the overall mean F0 difference between DDS and ADS sound stimuli being eliminated. This finding supports our hypothesis that the remaining averaged but still representative acoustic, temporal and paralinguistic features of the given acoustic stimuli were sufficient to

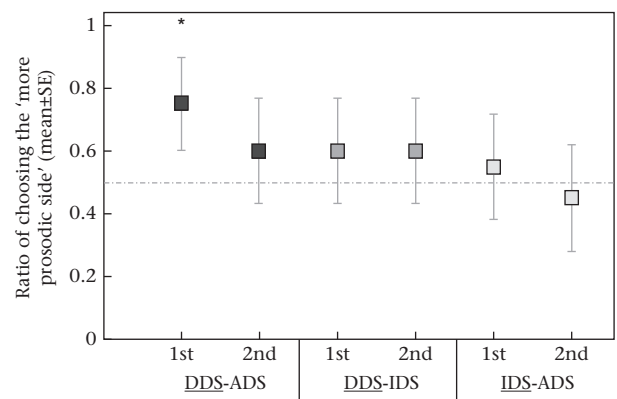


Figure 3. Probability of dogs choosing the ‘more prosodic’ stimulus in the first and second test trials in the three experimental conditions. The sound stimuli were dog-directed (DDS), infant-directed (IDS) or adult-directed (ADS). The horizontal line represents chance level (0.5). Underlines indicate the ‘more prosodic’ stimulus in the stimuli pairs in each condition. * $P < 0.05$, one-sample binomial test, $N = 20$ /condition.

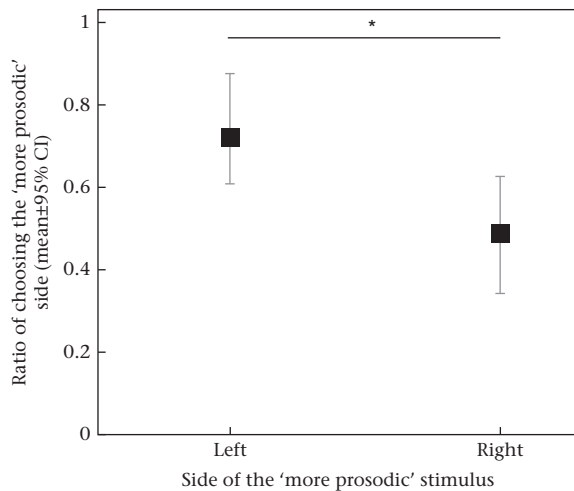


Figure 4. Probability of dogs choosing the 'more prosodic' stimulus when it was presented on the left or on the right side. * $P < 0.05$.

elicit a preference towards DDS but not towards IDS or ADS. This also suggests that DDS, without a higher overall F0 mean, remained distinguishable from ADS but not from IDS, which further confirms the widely reported phenomenon that DDS and IDS share numerous prosodic features (e.g. Burnham et al., 2002; Gergely et al., 2017; Jeannin et al., 2017). The lack of DDS preference in the second trial of the DDS–ADS condition, however, suggests that dogs' choices may be highly influenced by various factors: the trial number, the stimulus location (left/right) and the order of stimulus presentation in the two-way choice design.

Our experimental design, where only a single representative stimulus from a class of stimuli was used to test hypotheses about the whole class, raises a potential concern about pseudoreplication (Hurlbert, 1984). Namely, in studies using this particular design it is difficult to eliminate the possibility that some task-irrelevant stimulus features (e.g. any accidental attributes belonging to the sample stimulus of a particular addressee; dog/infant/adult) have an effect on the subjects' behaviour (Kroodsma, Byers, Goodale, Johnson, & Liu, 2001).

Our experiment was replicated for subjects (i.e. any two dogs in the same experimental group did not share more similar environmental conditions than any two dogs from different groups) but not for playback stimuli. One may therefore assume that if one 'irrelevant' detail of intonation, or noise etc., that renders the dogs highly responsive was included by chance in one of the playback stimuli, then the results could have been driven by that stimulus characteristic, and this limits the generalizability of our findings. Although an obvious solution to this problem is the use of multiple

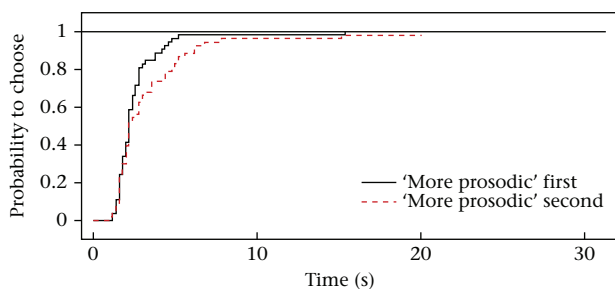


Figure 5. Probability over time of dogs choosing one of the tennis balls when hearing the 'more prosodic' stimulus first or second.

playback stimuli, many argue that pseudoreplication can also be reduced (at least to a certain extent) by using a composite stimulus that represents the average among several possible stimuli in a particular stimulus category (Patricelli, 2010; McGregor et al., 1992; Slabbekoorn, Ellers, & Smith, 2002).

In line with this suggestion, although using multiple stimuli would have been beneficial, our stimulus manipulation procedures (multistep modification of groups of natural DD, ID and AD sound stimuli) resulted in artificial stimuli that can be considered (at least to a certain extent) as representative of their respective broader categories (infant- dog- and adult-directed speech). Using 'synthetic templates' obtained by averaging the speech characteristics of different speakers (tone, intonation, pitch contour) can reduce the saliency of any random features irrelevant to the identification of dog- infant- or adult-directed sound stimuli.

Admittedly, however, there is no reason to assume that our composite stimuli would be fully representative of all stimuli in the DD, ID and AD classes. Thus, our study with unreplicated treatments provides less information than do those using multiple playback stimuli and these limitations cannot be overcome without further investigations. Concerning the potential role of the mean fundamental frequency, it has been suggested that higher overall F0 of DDS is crucial for dogs when discriminating DDS from ADS (Ben-Aderet et al., 2017). Others have claimed that the coexistence of specific lexical content and acoustic prosody is also essential to elicit DDS preference in adult dogs (Benjamin & Slocombe, 2018). Our study, however, points to the importance of other acoustic prosodic features beyond F0 mean in dog-directed verbal communication that seems to contribute to DDS identification in adult dogs. A previous study suggested that F0 range, intonation contour and harmonicity might be less important for attracting dogs' attention, while emphasizing that the coefficient of variation of the F0 and the intensity contour might play an important role (Jeannin et al., 2017). The three generated acoustic stimuli used in the present study did not allow for such correlation analysis between certain acoustic parameters of the stimulus and the dogs' responses. At the same time, our method for creating averaged DDS, ADS and IDS sounds has the potential to modify the acoustic parameters independently of one another and to investigate the effects of this particular prosodic feature on dogs' behaviour. In line with this, we will further investigate the effect of F0 variation and intensity contour modification with this method in future studies to clarify their role in DDS preference in dogs.

In line with our prediction and with the results of a previous study (Jeannin et al., 2017), dogs tended to respond similarly to IDS when it was paired with both ADS and DDS. It is reasonable to assume that dogs are not able to distinguish between IDS and DDS registers because of their similar acoustic and paralinguistic features (see Gergely et al., 2017). However, if dogs rely only on general prosodic differences in acoustic stimuli, they would be able to differentiate between IDS and ADS and would show some preference towards IDS (as it more resembles DDS). We cannot rule out the possibility that dogs were able to distinguish between ADS and IDS but still failed to show a preference for IDS because it was not directed to them. Note, however, that dogs were more willing to choose the 'more prosodic' side when it was on their left and their approach was faster when hearing the 'more prosodic' stimulus first. These results suggest that they could perceive the difference between ADS and IDS as well as between DDS and IDS. Jeannin et al. (2017) also found some evidence that dogs do distinguish between DDS and IDS, but their results were confounded by a strong stimulus order effect making it difficult to draw reliable conclusions. Dogs' ability to differentiate IDS, DDS and ADS needs further clarification and studies are also needed to investigate the

developmental and evolutionary aspects of looking/behavioural preferences towards DDS but not IDS.

Contrary to previous findings, in the present experiment we found no evidence for longer gazing at the location of 'more prosodic' sound stimuli in dogs (Benjamin & Slocombe, 2018; Jeannin et al., 2017). One plausible explanation would be that the DDS stimulus used in the present study was not as attention getting as an original and natural DDS used in these previous experiments due to its lowered mean F0 and the lack of lexical content (Benjamin & Slocombe, 2018; Jeannin et al., 2017). It has been suggested that dogs' attention and reaction are positively correlated with the overall F0 mean of the given sound (Ben-Aderet et al., 2017; Jeannin et al., 2017). In line with this assumption, our DDS stimulus with lowered F0 mean could have 'lost' its exaggerated attention-getting function. It has also been shown that the lexical content of dog-directed speech also matters for dogs at both neural and behavioural levels (Andics et al., 2016; Benjamin & Slocombe, 2018). It is likely, therefore, that sine waves are not as attention getting as natural DDS with relevant content. Alternatively, the lack of longer looking durations towards the 'more prosodic' stimulus location might be due to methodological differences between the present experiment and previous studies that showed increased gazing towards DDS stimuli. Jeannin et al. (2017) and Benjamin and Slocombe (2018) both used a protocol in which one or two female human experimenters were presented together with the acoustic stimulus (i.e. they were standing or sitting in front of the loudspeakers while avoiding eye contact with the subjects) to facilitate gazing towards the sound source. In the present study we used two tennis balls instead of live experimenters associated with stimulus locations, which might have resulted in shorter gazing durations in total and towards the 'more prosodic' stimulus location. By using target objects instead of a human in the present study we wanted to avoid the possibility that dogs associate the nonhuman speech-like sine wave sounds with the experimenter which could violate their expectation and affect their response. Note that the main findings of our study (significant preference for DDS over the ADS, a similar response to ADS and IDS, repetition and order effects) agree with the results of Jeannin et al. (2017); we can therefore assume that DDS preference over ADS is a more general phenomenon in dogs that can occur in various contexts and tasks.

The finding that choice latencies were faster when the 'more prosodic' stimulus was presented first is consistent with the hypothesis that dogs are able to perceive differences between DDS and ADS and also between IDS and ADS. We may assume that similarly to motherese for infants (e.g. Fernald, 1985), the 'more prosodic' stimuli (i.e. DDS and IDS) for dogs are more salient than ADS, and thus these stimuli have the potential to increase levels of arousal leading to faster approach.

Interestingly, our results also showed that dogs were more willing to choose the ball at the location of the 'more prosodic' sound source when it was on their left than when it was on their right. It is widely accepted that hemispheric lateralization in humans can cause such left-side bias when hearing prosody, as emotional processing shows a strong right-hemispheric dominance in adults (e.g. Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Seydell-Greenwald, Chambers, Ferrara, & Newport, 2020). Similar hemispheric asymmetry has been shown at both the behavioural and neural levels in dogs and this finding suggests a more ancient hemispheric specialization for acoustic and visual prosody processing (e.g. Racca, Guo, Meints, & Mills, 2012; Siniscalchi, Quaranta, & Rogers, 2008; Siniscalchi, Sasso, Pepe, Vallortigara, & Quaranta, 2010). Studies on lateralized visual/auditory behaviour typically apply the so-called head orienting (or dichotic listening) paradigm in which two stimuli sources are placed on the subjects' left and right (e.g. Gil-Da-Costa & Hauser,

2006; Ratcliffe & Reby, 2014). Given that auditory stimuli entering the right and left ears are processed mainly in the contralateral hemisphere, a right-ear advantage (right turn) reflects left-hemispheric dominance and a left-ear advantage (left turn) reflects right-hemispheric specialization (e.g. Grimshaw, Kwasny, Covell, & Johnson, 2003).

In light of this, we may assume that dogs are able to perceive the relative prosodic salience of these manipulated human speech stimuli (i.e. DDS > IDS > ADS), and that they tend to show a right-hemispheric predominance when processing it. This is surprising considering the equalized overall mean F0 of the DDS, IDS and ADS stimuli in the present experiment and further confirms the importance of acoustic, temporal and paralinguistic parameters other than F0 mean in dogs' prosody perception and processing.

Another notable aspect of the present study is that we used a novel method for creating general dog-, infant- and adult-directed acoustic stimuli, as the same sentence spoken by multiple female speakers was merged into a single audio clip. Previous experiments that aimed to study DDS preference in dogs or IDS preference in infants presented a single word or sentence or multiple sentences spoken by one female speaker to the subject, and tried to control for the speaker's identity by presenting different speaker voices ($N = 2-30$) to different subjects (e.g. Ben-Aderet et al., 2017; Benjamin & Slocombe, 2018; Fernald & Kuhl, 1987; Jeannin et al., 2017). We believe that our method provides a more effective control for the speaker's identity as individual features can be eliminated, while general characteristics of the speech register can be preserved. Moreover, this method allows systematic manipulation of acoustic, paralinguistic and temporal features of a given stimulus. To avoid pseudoreplication (e.g. Hurlbert, 1984), however, it would be feasible to create and use a set of composite (averaged) stimuli in future studies.

Acknowledgments

We are grateful to Sára Szabó and Krisztina Hegedűs-Kovács for their assistance in the data acquisition. This research received funding from the National Research Development and Innovation Office (PD121038. K128448) and the Premium Postdoctoral Scholarship of the Office for Research Groups Attached to Universities and Other Institutions of the Hungarian Academy of Sciences [460002].

Supplementary Material

Supplementary material associated with this article can be found online at <https://doi.org/10.1016/j.anbehav.2021.04.008>.

References

- Andics, A., Gábor, A., Gácsi, M., Faragó, T., Szabó, D., & Miklósi. (2016). Neural mechanisms for lexical processing in dogs. *Science*, 353(6303), 1030–1032. <https://doi.org/10.1126/science.aaf3777>
- Ben-Aderet, T., Gallego-Abenza, M., Reby, D., & Mathévon, N. (2017). Dog-directed speech: Why do we use it and do dogs pay attention to it? *Proceedings of the Royal Society B: Biological Sciences*, 284(1846). <https://doi.org/10.1098/rspb.2016.2429>
- Benjamin, A., & Slocombe, K. (2018). 'Who's a good boy?!' Dogs prefer naturalistic dog-directed speech. *Animal Cognition*, 21(3), 353–364. <https://doi.org/10.1007/s10071-018-1172-4>
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, 296(5572), 1435. <https://doi.org/10.1126/science.1069587>
- Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, 20(4), 477–488. [https://doi.org/10.1016/S0163-6383\(97\)90037-0](https://doi.org/10.1016/S0163-6383(97)90037-0)
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth preference for infant-directed first month after birth. *Child*

- Development*, 61(5), 1584–1595. <https://doi.org/10.1111/j.1467-8624.1990.tb02885.x>
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2), 181–195. [https://doi.org/10.1016/S0163-6383\(85\)80005-9](https://doi.org/10.1016/S0163-6383(85)80005-9)
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–1510. <https://doi.org/10.2307/1130938>
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279–293. [https://doi.org/10.1016/0163-6383\(87\)90017-8](https://doi.org/10.1016/0163-6383(87)90017-8)
- 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. *Scientific Reports*, 7, 13739. <https://doi.org/10.1038/s41598-017-13883-2>
- Gil-Da-Costa, R., & Hauser, M. D. (2006). Vervet monkeys and humans show brain asymmetries for processing conspecific vocalizations, but with opposite patterns of laterality. *Proceedings of the Royal Society B: Biological Sciences*, 273(1599), 2313–2318. <https://doi.org/10.1098/rspb.2006.3580>
- Grimshaw, G. M., Kwasny, K. M., Covell, E., & Johnson, R. A. (2003). The dynamic nature of language lateralization: Effects of lexical and prosodic factors. *Neuropsychologia*, 41(8), 1008–1019. [https://doi.org/10.1016/S0028-3932\(02\)00315-9](https://doi.org/10.1016/S0028-3932(02)00315-9)
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2), 187–211.
- Jeannin, S., Gilbert, C., Amy, M., & Leboucher, G. (2017). Pet-directed speech draws adult dogs' attention more efficiently than Adult-directed speech. *Scientific Reports*, 7, 4980. <https://doi.org/10.1038/s41598-017-04671-z>
- Kroodtsma, D. E., Byers, B. E., Goodale, E., Johnson, S., & Liu, W. C. (2001). Pseudoreplication in playback experiments, revisited a decade later. *Animal Behaviour*, 61(5), 1029–1033. <https://doi.org/10.1006/anbe.2000.1676>
- Liu Tsao, K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language*, 36(4), 909–922. <https://doi.org/10.1097/MPG.0b013e3181a15ae8>
- McGregor, P. K., Catchpole, C. K., Dabelsteen, T., Falls, B. J., Fusani, L., Gerhardt, C. H., et al. (1992). Design of playback experiments: The Thornbridge Hall NATO ARW consensus. In P. K. McGregor (Ed.), *Playback and studies of animal communication. NATO ASI Series* (pp. 1–9). Boston, MA: Springer.
- Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, 41(10), 1410–1421. [https://doi.org/10.1016/S0028-3932\(03\)00017-4](https://doi.org/10.1016/S0028-3932(03)00017-4)
- Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., & Nakamura, K. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage*, 59(2), 1735–1744. <https://doi.org/10.1016/j.neuroimage.2011.07.093>
- Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2020). The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. *Developmental Science*, 24(1), Article e12997. <https://doi.org/10.1111/desc.12997>
- Patricelli, G. L. (2010). Robotics in the study of animal behavior. In J. Breed, & M. Moore (Eds.), *Encyclopedia of animal behavior* (pp. 91–99). Westport, CT: Greenwood Press.
- Pisanski, K., Jones, B. C., Fink, B., O'Connor, J. J. M., DeBruine, L. M., Röder, S., et al. (2016). Voice parameters predict sex-specific body morphology in men and women. *Animal Behaviour*, 112, 13–22. <https://doi.org/10.1016/j.anbehav.2015.11.008>
- Racca, A., Guo, K., Meints, K., & Mills, D. S. (2012). Reading faces: Differential lateral gaze bias in processing canine and human facial expressions in dogs and 4-year-old children. *PLoS One*, 7(4), Article e36076. <https://doi.org/10.1371/journal.pone.0036076>
- Ratcliffe, V. F., & Reby, D. (2014). Orienting asymmetries in dogs' responses to different communicatory components of human speech. *Current Biology*, 24(24), 2908–2912. <https://doi.org/10.1016/j.cub.2014.10.030>
- Seydell-Greenwald, A., Chambers, C. E., Ferrara, K., & Newport, E. L. (2020). What you say versus how you say it: Comparing sentence comprehension and emotional prosody processing using fMRI. *NeuroImage*, 209, 116509. <https://doi.org/10.1016/j.neuroimage.2019.116509>
- Siniscalchi, M., Quaranta, A., & Rogers, L. J. (2008). Hemispheric specialization in dogs for processing different acoustic stimuli. *PLoS One*, 3(10), e3349. <https://doi.org/10.1371/journal.pone.0003349>
- Siniscalchi, M., Sasso, R., Pepe, A. M., Vallortigara, G., & Quaranta, A. (2010). Dogs turn left to emotional stimuli. *Behavioural Brain Research*, 208(2), 516–521. <https://doi.org/10.1016/j.bbr.2009.12.042>
- Slabbekoorn, H., Eilers, J., & Smith, T. B. (2002). Birdsong and sound transmission: The benefits of reverberations. *The Condor: Ornithological Applications*, 104(3), 564–573.
- Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America*, 128(1), 389–400. <https://doi.org/10.1121/1.3419786>
- Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, 18(5), 727–735. <https://doi.org/10.1037/0012-1649.18.5.727>
- Sulpizio, S., Hirokazi, D., Bornstein, M. H., Joy, C., Gianluca, E., & Kazuyuki, S. (2018). fNIRS reveals enhanced brain activation to female (versus male) infant directed speech (relative to adult directed speech) in Young Human Infants. *Infant Behavior and Development*, 52, 89–96. <https://doi.org/10.1103/PhysRevLett.119.016101>
- Titze, I. R. (2000). *Principles of voice production*. Iowa City, IA: National Center for Voice and Speech.
- Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyper-articulation in Parrot-, dog- and infant-directed speech. *Anthrozoös: A Multidisciplinary Journal of The Interactions of People & Animals*, 26(3), 373–380. <https://doi.org/10.2752/175303713X13697429463592>