Journal of Child Language (2025), 1–17
doi:10.1017/S0305000924000552

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

More than just a happy talk? Evidence for functional pitch and utterance length modifications in infant-, spouse-, and dog-directed communication

Édua Koós-Hutás[1,2], Shanjida Afrin[3], Alexandra Barbara Kovács[2,3], Tamás Faragó[3,4], Lőrinc András Filep[2], József Topál[2] and Anna Gergely[2]

[1]Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary, [2]ELTE-HUN-REN NAP Comparative Ethology research group, Research Centre for Natural Sciences, Institute of Cognitive Neuroscience and Psychology, Budapest, Hungary, [3]Institute of Biology, Department of Ethology, Eötvös Loránd University, Budapest, Hungary and [4]Department of Ethology, Neuroethology of Communication Lab, Budapest, Hungary

Corresponding author: Édua Koós-Hutás; Email: koos.edua@ppk.elte.hu

(Received 13 November 2023; revised 02 September 2024; accepted 09 September 2024)

Abstract

By comparing infant-directed speech to spouse- and dog-directed talk, we aimed to investigate how pitch and utterance length are modulated by speakers considering the speech context and the partner's expected needs and capabilities. We found that mean pitch was modulated in line with the partner's attentional needs, while pitch range and utterance length were modulated according to the partner's expected linguistic competence. In a situation with a nursery rhyme, speakers used the highest pitch and widest pitch range with all partners suggesting that infant-directed context greatly influences these acoustic features. Recent findings showed that these speakers expressed more intense positive emotions towards their infants and spouses than towards their dogs. Our results revealed different patterns, leading us to conclude that these acoustic features are not simple by-products of emotional speech. Instead, they are dynamically and functionally used in accordance with the speech context and the audience's expected needs and capabilities.

Keywords: prosody; infant-directed speech; language tutoring; parentese

Introduction

Usually, speakers change their speech style based on their listener by adjusting various acoustic characteristics that are associated with prosody, including mean overall pitch (fundamental frequency (*f0)* mean) and other pitch-related features (*f0* range, variability, contour, etc.) (e.g. Falk, 2004; Saint-Georges et al., 2013). While talking to infants, caregivers tend to use higher overall pitch, wider pitch range, specific pitch contours, and longer utterances (e.g. Burnham et al., 2002; Golinkoff et al., 2015). There is ample

evidence that the characteristics of infant-directed speech prosody serve multiple functions. These functions include capturing and maintaining the infant's attention, strengthening the bond between the infant and the caregiver through enhanced positive interactions, facilitating language acquisition, expressing emotions, and conveying information about the speaker's intentions and identity. As a result, infant-directed speech plays an essential role in the healthy emotional and cognitive development of children (for a review, see Soderstrom, 2007). In the past decades, there has been growing interest in more systematic and controlled investigations, which have the potential to reveal more exact functions and related acoustic features in infant-directed speech prosody. In the present study, we focused on the potential functions of two pitch-related characteristics (*f0* mean and range) and one utterance length-related feature (call length) of infant-directed speech prosody.

Effect of situation

One approach is to investigate and compare infant-directed speech prosody in and between different situations and contexts. With this method, it has been shown that various pitch characteristics can play distinct functions and roles during tutoring interactions with preverbal infants. More precisely, specific large pitch contours in infant-directed speech (which manifests in a wider *f0* range) have the potential to facilitate word segmentation and, thus, language acquisition (e.g. Thiessen et al., 2005; Trainor & Desjardins, 2002). By contrast, it has been suggested that a higher overall pitch (i.e. *f0* mean) not only does not facilitate but actually also impedes word segmentation. Simultaneously, it plays an essential role in capturing and controlling infants' attention and expressing emotions (e.g. Cooper & Aslin, 1994; Trainor & Desjardins, 2002). It has also been suggested that tutoring and playing situations involving objects contain less exaggerated prosody (i.e. lower *f0* mean and smaller *f0* range) to effectively divide infants' attention between the object and the speaker (e.g. Gergely et al., 2017; Gogate et al., 2006).

The relevancy (i.e. infant directedness) and naturalness (i.e. fixed sentences or text reading versus free speech) of the given situation also affect the pitch characteristics of prosody. When a specific text, such as a story from a book, had to be read to children, speakers used lower *f0* mean and smaller *f0* range compared to situations where they were allowed to speak freely to the infant (e.g. Shute & Wheldall, 1999; Gergely et al., 2017). At the same time, fixed sentences that are pronounced rhythmically and melodically and have typical infant-directed content (e.g. rhymes and playsongs) seem to have distinctive, intense, and consistent acoustic prosody with heightened *f0* mean (e.g., Falk & Audibert, 2021).

Effect of the partners' needs and capacities

Another feasible approach to studying functions of prosody and related acoustic features is to compare them across different partners with varying emotional needs and cognitive capacities. Using such a comparative method, it has been revealed that people tend to employ strikingly similar acoustics, including higher *f0* mean and wider *f0* range when talking to infants and pets, which differ significantly from the speech towards unfamiliar adults (e.g. Hirsh-Pasek & Treiman, 1982; Burnham et al., 2002; Gergely et al., 2017). It has been suggested that one basic function of such exaggerated prosody is to evoke and maintain the attention of partners with limited linguistic competence, whether conspecific or heterospecific (e.g. Hirsh-Pasek & Treiman, 1982; Burnham et al., 2002; Gergely et al.,

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

2017). Besides the acoustic similarities, there is also evidence that the given context and the naturalness of the situation similarly influence the *f0* mean and range of infant- and dog-directed speech (e.g. Gergely et al., 2017).

This comparative framework has also revealed a relationship between utterance lengthening (i.e. vowel hyperarticulation) and the linguistic competence of the intended addressee: speakers used the longest vowels towards infants (i.e. future speakers) than towards parrots (i.e. expected future speakers), but not towards dogs or cats (i.e. non-speakers; e.g. Burnham et al., 2002; Gergely et al., 2017; Xu et al., 2013). The aforementioned results supported the language tutoring function of utterance lengthening and provided evidence that, similarly to pitch characteristics, speakers adjust these parameters as well to their audience's expected needs and capacities.

Conveying positive emotions, expressing affection, and strengthening attachment are listed among the most important functions of infant-directed prosody, to which heightened and wider-ranged *f0* contributes greatly (e.g. Fernald, 1992; Trainor et al., 2000). Moreover, it has been suggested that the striking acoustic differences between adult- and infant-directed prosody are by-products of speakers' emotional expressions when interacting with infants and inhibited when talking to adults (Trainor et al., 2000). Facial expressions accompanied by infant- and adult-directed acoustic prosody seem to support this notion, as more exaggerated facial expressions are displayed towards infants than towards adult partners (e.g. Chong et al., 2003; Gergely et al., 2023). It is important to note, however, that in the aforementioned studies, prosody towards one's own infant was compared to speech prosody towards a nice but unfamiliar adult partner (i.e. experimenters). As attachment and personal relationships between the interactants greatly impact speakers' emotions and speech prosody (e.g. Bombar & Littig, 1996), the feasibility of comparing speech prosody towards unfamiliar adults and own infants has been questioned (Trainor et al., 2000).

Xu and co-workers (2013) used the same unfamiliar partners (adult, dog, or parrot) with all female speakers in their study and provided evidence that acoustic differences between adult- and pet-directed speech are still evident when familiarity between conditions is equalized (Xu et al., 2013). In a recent study, Koós-Hutás and co-workers (2024) compared facial emotional expressions and emotional states of female and male speakers when interacting with their 6- to 18-month-old infants, their spouses, and their family dog. Contrary to previous findings with unfamiliar adult partners, speakers in this study showed similarly intense emotions and related facial expressions during infant- and adult (i.e. spouse)-directed conditions (Gergely et al., 2023; Koós-Hutás et al., 2024). These results highlight the importance of taking personal relationships into account between the interactants (Trainor et al., 2000). It is also important to note that, in this study, speakers used less intense and less positive facial expressions with their family dogs than with their infants and spouses suggesting that facial expressions might follow different dynamics and have different functions than pitch characteristics, which speakers use similarly with dogs and infants (Hirsh-Pasek & Treiman, 1982; Gergely et al., 2017; Koós-Hutás et al., 2024).

Effect of the speakers' sex

According to the current state of the literature, acoustic features as well as utterance length-related properties of infant-directed speech are more similar than different among women and men (for a review, see Ferjan Ramírez, 2022). There is ample evidence that both sexes use higher pitch during infant-directed speech than during adult-directed speech

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

(e.g. Niwano & Sugai, 2003; Gergely et al., 2017; Weirich & Simpson, 2019). Pitch range, on the other hand, presents a more variable picture of how sex differences are manifested in infant- and adult-directed conditions. Several studies have reported wider pitch range in female speakers than in male speakers during parent–infant interactions in various contexts and languages, including spontaneous and read speech situations (e.g. Fernald et al., 1989; Gergely et al., 2017). However, other studies have found no sex differences in infant-directed pitch range (e.g. Shute & Wheldall, 1999; Niwano & Sugai, 2003) or have shown that male speakers use a wider range than female speakers (e.g., Warren-Leubecker & Bohannon, 1984). When it comes to pet-directed speech, there is evidence that both sexes use similarly heightened pitch and wide pitch range when talking to dogs as opposed to adults, but similar to that of infant-directed speech (Gergely et al., 2017). Moreover, both sexes hyperarticulate their vowels with infants, but not with dogs and unfamiliar adults (e.g. Burnham et al., 2002; Gergely et al., 2017).

Aims and hypotheses

In the present study, we aimed to investigate the functions of two pitch-related parameters (*f0* mean and range) of infant-directed acoustic prosody by comparing them across different situations and partners in both women and men. To achieve this, we analysed speech samples from our recently published comparative study (Koós-Hutás et al., 2024), in which female and male speakers interacted with their own infants (infant-directed condition), own spouses (adult-directed condition), and own family dogs (dog-directed condition) during two free speech situations (attention getting and language tutoring) and one fixed sentences situation with a nursery rhyme (fixed sentences). In addition to *f0* mean and range, we also aimed to study one utterance length-related parameter (call length) during the language tutoring situation, to examine whether speakers adjust their uttering in line with the partners' expected linguistic competence.

Our first research question was as follows: (1) whether and how different speech situations affect the speakers' mean pitch and pitch range towards their infants, spouses, and dogs. Heightened *f0* mean proved to be crucial for capturing and maintaining the attention of partners with limited linguistic competence (i.e. infants and dogs; e.g. Fernald & Kuhl, 1987; Jeannin et al., 2017). However, a heightened *f0* mean might impede word segmentation, while a wider *f0* range has the potential to facilitate language acquisition (e.g. Trainor & Desjardins, 2002). We hypothesized, therefore, that the attention-getting situation, in which speakers were instructed to get and maintain the focus of their partners on themselves, would evoke higher *f0* mean when speaking to infants and dogs compared to adults. Additionally, we predicted that speakers would use a lower *f0* mean and wider *f0* range when talking to infants compared to dogs during the language tutoring situation. Concerning the fixed sentences situation, in which speakers were instructed to tell three everyday-like sentences along with a nursery rhyme to the partners, we could predict two different outcomes based on the literature. There is evidence that speakers use less exaggerated prosody with their partners during less naturalistic and more restricted situations (e.g., Gergely et al., 2017; Jürgens et al., 2011) which suggests lower *f0* mean and smaller *f0* range in this situation compared to the two free speech situations. On the other hand, it has also been shown that rhythmic and melodic speech and the infant directedness of a speech affect prosody and can evoke intense acoustics from the speakers (e.g. Falk & Audibert, 2021). Therefore, it is also possible that the fixed sentences situation with a nursery rhyme will evoke similar or even more exaggerated prosody with a higher *f0* mean and wider *f0* range, irrespective of the type of the partner, compared to the free speech situations.

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

The second research question of the present study was as follows: (2) whether and how speakers adjust mean pitch, pitch range, and utterance length according to their partners' expected language competence. If such adjustments occur, we would expect a higher mean *f0* and a wider *f0* range when addressing partners with developing linguistic skills (i.e. infants) or limited linguistic skills (i.e. dogs) compared to fully competent speakers (c.f. Burnham et al., 2002; Gergely et al., 2017). Based on the results of previous studies on hyperarticulation and acoustics (e.g. Trainor & Desjardins, 2002; Burnham et al., 2002; Gergely et al., 2017; Xu et al., 2013), we may expect that speakers will use longer utterances (i.e. call length), lower *f0* mean, and wider *f0* range to facilitate word segmentation for potential speakers (i.e. infant) when uttering a to-be-thought object label (i.e. language tutoring situation). However, shorter utterances (i.e. call length), higher *f0* mean, and smaller *f0* range are expected when speakers are uttering it to non-speakers (i.e. dogs; Burnham et al., 2002; Gergely et al., 2017; Xu et al., 2013). Speakers are expected to use no speech modifications to enhance word segmentation when interacting with equally competent speakers (i.e. their spouses, e.g. Burnham et al., 2002; Gergely et al., 2017).

Alternatively, it is also possible that speakers' emotions play a more significant role in regulating speech prosody than the audience's needs and capabilities. Recently, these speakers' facial expressions and related emotional content were analysed and showed that both female and male speakers in all examined situations used more frequent and intense happy emotions when interacting with their infants and spouses than with their dogs (Koós-Hutás et al., 2024). We can hypothesize that the acoustics of the accompanied speech will follow this emotional pattern of the speakers, and as a "by-product" of happy speech, we can predict heightened and wider-ranged *f0* when interacting with the spouses and infants than with the dogs (e.g. Fernald, 1992; Trainor et al., 2000).

The third research question we aimed to study was as follows: (3) whether and how speakers' sex affects the two pitch-related and one utterance length-related parameters of their speech. Based on the literature, aforementioned hypotheses, and predictions regarding *f0* mean, we expect similar patterns in female and male speakers (e.g. Niwano & Sugai, 2003; Gergely et al., 2017). However, a wider f0 range will likely be observed in female speakers compared to male speakers (Fernald et al., 1989; Gergely et al., 2017). According to previous results, we also expect female and male speakers to modulate their utterance length similarly (e.g. Burnham et al., 2002; Gergely et al., 2017).

Materials and methods

Ethics statement

This research was approved by the Human Research Ethics Committee (EPKEB) at the Hungarian Academy of Sciences (No. 2022-85). All parents gave their written consent to engage in the research in accordance with ethics approval, and all procedures were carried out in accordance with the relevant rules and regulations of the EPKEB and the applicable laws of Hungary.

Participants

Both parents from 22 families (N=44; 22 women and 22 men; mean age ± standard deviation [SD]: 34.6 ± 4.4 years; urban, heterosexual, and middle-class families) voluntarily participated in this research (Koós-Hutás et al., 2024). Each family had their own infant (6–18 months old; 10 girls and 12 boys; mean age ± SD: 10.2 ± 3.7 months) and a

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

pet dog that is at least 1 year old (13 female and 15 male dogs; mean age ± SD: 6 ± 3.7 years). All the parents were instructed to interact with their baby (infant-directed condition) and their family dog (dog-directed condition). If there were more than one dog in the family, the speakers had the liberty to interact with different dogs, choosing those with whom they felt most comfortable. During the adult-directed condition, they interacted with their spouses. All participants had Hungarian as their first language. Demographic details of the participating interactants are reported in the supplementary material (Table S1).

Procedure

Data collection took place at the participants' homes in the presence of two experimenters. One of them managed the technical equipment required for the recording, while the other supervised the entire process. Before beginning, each parent signed an informed consent form. After that, each mother and father were recorded individually while interacting with their own infant, dog, and spouse in a within-subject design. Speakers were instructed to occupy seats about 30 centimetres away from the addressee at eye level or lower to avoid data loss of the speaker's face by gazing down (see Figure 1; Koós-Hutás et al., 2024). Leaning over or touching the addressee in certain circumstances was not strictly forbidden, but the speakers were encouraged to try to maintain their position throughout the interaction. Adult partners (i.e. spouses) were instructed to maintain a sit position during the experiment, and dogs were placed in a sit or down position at the same spot, while infants were sitting in a baby chair or the spouse's lap or the experimenter's lap during the interactions (see Figure 1).

Speech interactions were recorded in three different situations – attention getting, language tutoring, and fixed sentences – using the same microphone (Zoom F2 recorder with LMF-2 Lavalier microport). Smartphones were also used during the study to record data for a separate analysis, which was reported in another study (Koós-Hutás et al., 2024). Participants were told to engage in natural conversation with the addressees during each recording phase, which consisted of three situations. The order of situations and conditions was counterbalanced across participants.

Attention-getting situation (1 minute)

Participants were told to capture the addressee's attention and maintain his/her attentional focus (preferably by maintaining eye contact) for one minute. We aimed to observe

Figure 1. Experimental arrangement. (a) Dog-directed condition, (b) adult (spouse)-directed condition, and (c) infant-directed condition.

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

how the speaker naturally manages to maintain the addressee's attention, so we did not provide specific instructions to the speakers on how to complete the tasks.

Language tutoring situation (1+1 minutes)

During this situation, speakers were instructed to teach an object–label association to their partners (presentation phase), and then, the partner was asked to select the labelled object (two-way choice task). To do so, the experimenter chose randomly two objects out of five, all of which were novel to the partners (see Figure 2). One object was randomly assigned as a target object and the other one as a non-target object. Then, the experimenter randomly selected one of the predetermined three artificial words ("danidu," "burida," and "zibula") and asked the speaker to label the target object using this word while interacting with his/her baby, dog, or spouse. When creating the words for object labels, we aimed to use novel words without meaning that interactants had never heard before. Note that all labels were required to contain the three syllables necessary to draw vowel triangles (i.e. i, a, u) for future studies aiming to investigate hyperarticulation.

Language tutoring – presentation phase (1 minute). The speakers' task was to associate artificial labels with the target object while holding both the target and non-target objects in their hands. Speakers were instructed to use only demonstrative words such as "this," "that," "thing," and "something" when referring to the non-target object. They were told to talk about both the target and non-target objects separately for at least half a minute, using the predetermined label (referring to the target) and the demonstrative words (referring to the non-target) as frequently as it is possible (for a similar method, see Woodward et al., 1994). The addressee was not allowed to touch the objects during this phase.

Language tutoring – two-way choice task phase (1 minute). After about a minute, the speaker moved on to the second phase and encouraged the addressee to select the target object with these words: "Which one is the danidu/burida/zibula?". During this phase, speakers were instructed to hold the two objects still at an equal distance (at arm's length) from the addressee. If needed, speakers were allowed to encourage the partner verbally to choose without moving the objects. After choosing an object, the addressee was allowed to touch and explore the chosen object, and the speaker was allowed to praise the partner. Then, the speaker kindly asked for the object back from the partner, switched the position of the target and non-target objects in her/his hands, and repeated the whole "choosing" procedure once more.

Figure 2. Set of potential target and non-target objects used in the language tutoring situation.

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

Fixed sentences situation (1 minute)

Participants were instructed to recite a nursery rhyme and three previously specified sentences to the addressee. The fixed three sentences were as follows: (#1) *Nézd csak, milyen szép idő van odakint!* (in English: Just look! What nice weather!), (#2) *Akarsz sétalni egyet?* (in English: Do you want to go for a walk?), and (#3) *Úgy látom, unatkozol. Nem csinálunk valami mást?* (in English: You seem really bored. Shouldn't we do something else?).

Apart from the three fixed sentences, speakers were also asked to recite the following well-known Hungarian nursery rhyme: *Cini-cini muzsika; táncol a kis Zsuzsika; jobbra dől, balra dől; tücsök koma hegedül* (in English: "Cini-cini music plays; little Susan dances away; leaning to the right, leaning to the left; the cricket buddy plays the fiddle").

Data analysis

Acoustic analysis

We used acoustic data (i.e. the audio file recorded by the microport) from our recent study in which only the facial prosodic features of the speakers were analysed (Koós-Hutás et al., 2024). The analysis of the acoustic recordings from all three situations was done in line with Gergely et al., 2017, with the help of the Praat software (version 6.0.05; Boersma & Weenink, 2021). It is important to note that for the analysis we used only recordings of the Zoom microport and not the smartphones. At first, we used a semi-automatic script to annotate the recordings, defining and labelling pauses and calls and excluding background sounds. We applied a call-based approach for our analyses similar to Gergely et al., 2017. One call, in terms of bioacoustics, can be considered as a functional unit in the speech stream intonation contour which usually contains one voiced sound. Calls are separated by pauses, breathtaking, and unvoiced sounds, similarly to utterance units. The baseline search range was defined between 75 Hz and 500 Hz, and before the pitch extraction, the coder checked visually the detection of the pitch contour for halving and doubling errors and modified the range if it was necessary. This way we could ensure the minimal level of artefacts in the measurements and we could also exclude intermittent vocalizations as well as remaining background noises from the sample. Then, we exported the following acoustic characteristics of each call from the programme:

f0 mean: It refers to the mean of the fundamental frequency (*f0*, perceived as pitch) of each call (40148 calls in total, 13620 in adult-directed, 13274 in dog-directed, and 13254 in infant-directed conditions). The analysis was performed using Praat's built-in cross-correlation-based pitch extraction method.

f0 range: The Praat software's built-in function was used to calculate each call's *f0* range by subtracting *f0* minimum from *f0* maximum.

Call length: The Praat software's built-in function was used to analyse the call length of the object labels. This analysis aimed to investigate whether speakers uttered the label differently when talking to infants, dogs, and adults. When labels were not isolated, we manually separated them in Praat software by using tiers, ensuring that all labels (i.e. danidu/burida/zibula) were analysed as a single continuous call (2112 calls/labels in total).

Statistical analysis

RStudio (https://www.rstudio.com/) was used for the statistical analysis (R version 4.2.3 using RStudio 2023.06.0+421, R Core Team 2023). To analyse *f0* mean and range, we used

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

generalized linear mixed models (nlme and lme4 package and glmer and lme functions; Bates et al., 2015; Pinheiro & Bates, 2000) with the Akaike information criterion (AIC)-based backwards elimination (MASS package and drop1 function; Venables & Ripley, 2002) to find parsimonious models. Due to the anatomy-based difference in *f0* mean of women and men (Titze, 1989), *f0* mean of female and male speakers was analysed with separate models for the whole dataset and for the object label analysis. As the data distribution was skewed towards low values, we normalized them with log transformation. Also, as fixed sentences situation had lower variance, we controlled for heteroscedasticity in these models by adding situation-dependent weights to the model. In *f0* mean models, for the whole dataset, condition (infant-, adult-, and dog-directed), situation (attention getting, language tutoring, and fixed sentences), and their interaction were included as fixed effects. For *f0* range and call length analysis, female and male speakers were included in the same model; therefore, the effect of sex (female and male) and all two- and three-way interactions with condition (*f0* range and call length) and situation (*f0* range) were included. In object label models (*f0* mean, *f0* range, and call length variables), condition, sex, and their interaction were included. First, we included speaker identity number (ID) and family ID to the models as random intercepts (speaker nested in family) to control for dependence and repeated measurements. After comparing model performance (compare performance function) and checking the explained variance, family ID was dropped out as it explained no variance, and only speaker ID was included as a random intercept in all final models. For post hoc pairwise comparisons, we used the Tukey method (emmeans package; Lenth, 2023).

Results

First, we will present the significant interactions and main effects (i.e. situation, condition, and speakers' sex) for all analysed prosodic features (i.e. *f0* mean and range, call length). Then, we will present the post hoc analysis and pairwise comparisons according to the research questions (for summary, see Table 1).

Significant interactions and main effects

According to the *f0* mean (all calls), model selection showed a significant interaction effect of condition × situation in both female (*LRT*: χ^2_4=133.93, p<0.001) and male (*LRT*: χ^2_4=98.885, p<0.001) speakers. According to the *f0* range (all calls), model selection showed a significant interaction effect of condition × situation (*LRT*: χ^2_4=16.04, p<0.001), speakers' sex × condition (*LRT*: χ^2_2=8.03, p=0.018), and speakers' sex × situation (*LRT*: χ^2_2=10.28, p=0.006). When it comes to the object labels, the model selection of *f0* mean (labels) showed a significant main effect of condition in both female

Table 1. Summary of the acoustic features of object labels during the language tutoring situation in all three conditions

	f0 mean	*f0* range	Call length
Infant-directed speech	High	Wide (females)	Long
Dog-directed speech	High	Narrow	Short
Adult-directed speech	Low	Narrow	Long

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

(*LRT*: χ^2_2=98.83, p<0.001) and male (*LRT*: χ^2_2=36.71, p<0.001) speakers. In object labels, the model selection also showed a significant interaction effect of speakers' sex × condition both for *f0* range (labels, *LRT*: χ^2_2=6.72, p=0.035) and for call length (labels, *LRT*: χ^2_2=20.86, p=0.035).

Effect of speech situation

Speakers used similarly high pitch during fixed sentences and attention-getting situations when interacting with their infants (p>0.05), but a lower *f0* mean was observed during the language tutoring situation in both male and female speakers (all p<0.05; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). Speakers used the highest *f0* mean during the fixed sentences situation (all p<0.05) and a similarly lower one in attention-getting and language tutoring situations when talking to their dogs (p>0.05; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). When interacting with their spouses, both sexes used the highest *f0* mean during fixed sentences, followed by language tutoring and finally during attention-getting situations (all p<0.05; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics).

Pairwise comparisons revealed general patterns of speech situation on speakers' *f0* range. The widest range was observed during the fixed sentences situation, followed by language tutoring and finally in attention-getting situation in both sexes across all three conditions (all p<0.05; see Figure 4 for summary, and see Table S3 and Figure S2 for detailed statistics).

Effect of the partners' linguistic competence

Pairwise comparisons of *f0* mean showed that both female and male speakers used a higher *f0* mean towards their infants and dogs than towards their spouses in all three

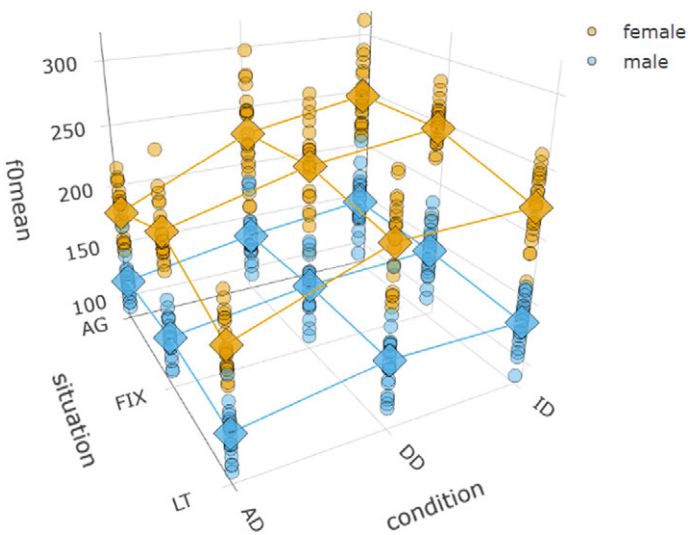

Figure 3. Fundamental frequency (*f0*) mean (Hz) of female and male speakers during situations across all three conditions.

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

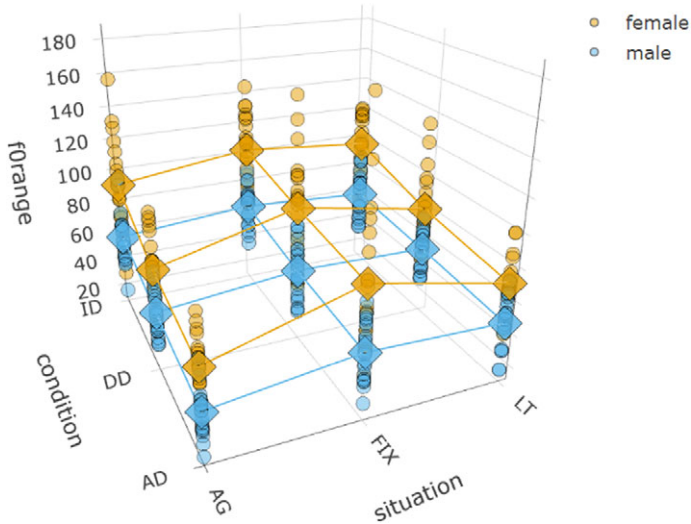

Figure 4. Fundamental frequency (*f0*) range (Hz) of female and male speakers during situations across all conditions.

situations (all p≤0.001; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). *F0* mean was similar towards infants and dogs in female speakers during the language tutoring situation and in male speakers during the fixed sentences situation (both p>0.05; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). However, the pattern of *f0* mean towards infants compared to dogs exhibited greater diversity. In the attention-getting situation, speakers from both sexes employed a higher *f0* mean towards infants than towards dogs (all p<0.05; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). In the language tutoring situation, male speakers used an even higher *f0* mean towards dogs than towards infants, while female speakers maintained a similar *f0* mean towards dog and infant partners during this situation (Figure 3; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics). During the fixed sentences situation, female speakers used a higher *f0* mean with infants than with dogs, while male speakers maintained a similar *f0* mean across infant-directed and dog-directed interactions in this situation (Figure 3; see Figure 3 for summary, and see Table S2 and Figure S1 for detailed statistics).

In both female and male speakers, the widest *f0* range was observed towards infants, then towards dogs, and finally towards adults in almost all situations. The only exception was detected in the fixed sentences situation, during which infant- and dog-directed speech contained a similar *f0* range (see Figure 4 for summary, and see Table S3 and Figure S2 for detailed statistics).

Pairwise comparisons showed that male speakers used the highest *f0* mean when uttering object labels towards their dogs, followed by their infants and finally towards their spouses (all p<0.05; see Table 1 for summary, and see Table S4 and Figure S3 for detailed statistics). At the same time, female speakers used similarly high *f0* mean when conveying object labels to their dogs and infants, while they also used a lower *f0* mean when conveying the object labels to their spouses (see Table 1 for summary, and see Table S4 and Figure S3 for detailed statistics).

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

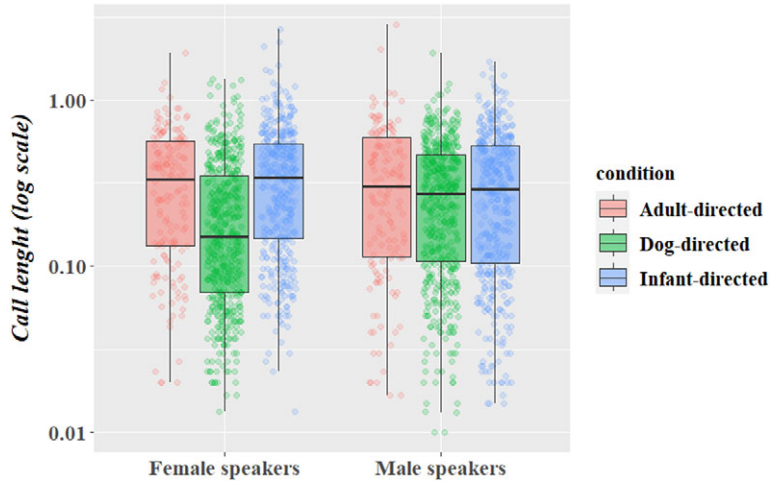

Figure 5. Call length of object labels of female and male speakers in all three conditions. Within the boxplots, the horizontal line represents the median, the box indicates the quartiles, the whiskers represent the range, and the dots represent the individual data points.

Pairwise comparisons also showed that female speakers used a wider *f0* range of object labels when speaking to infants compared to dogs or adults (all $p<0.05$). However, they used a similar range when addressing dogs and adults (all $p>0.05$; see Table 1 for summary, and see Table S4 and Figure S4 for detailed statistics). Additionally, male speakers used a similar range when conveying object labels to infants, dogs, and adults (all $p>0.05$, Figure 4; see Table 1 for summary and Table S4 for detailed statistics). Consistent with the *f0* range model results on the whole dataset, female speakers generally exhibited a wider *f0* range than males across all conditions (all $p>0.05$, Figure 4; see Table 1 for summary and Table S4 for detailed statistics).

We found that both sexes uttered the object label longer to their infants and their spouses than towards their dogs, while they used a similar call length towards their infants and their spouses (Figure 5; see Table 1 for summary and Table S4 for detailed statistics). Call length was similar between sexes in all conditions (all $p>0.05$, Figure 5; see Table 1 for summary and Table S4 for detailed statistics).

Effect of speakers' sex

In line with our hypothesis, pairwise comparisons revealed general patterns of the sex on speakers' *f0* range. Female speakers used a wider *f0* range than male speakers during all situations and across all conditions (all $p<0.05$; see Figure 4 for summary, and see Table S3 and Figure S2 for detailed statistics).

Discussion

In the present study, we investigated and compared two pitch-related parameters (*f0* mean and range) as well as one utterance length-related parameter (call length) of female and male speakers' speech during interactions with their own infants (infant-directed

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

speech), their own family dogs (dog-directed speech), and their spouses (adult-directed speech). These interactions were observed in two free speech situations (attention getting and language tutoring) and one fixed sentences situation with a nursery rhyme (fixed sentences). Our aim was to study whether and how the different situations, the partners' expected linguistic competence, the speakers' emotions, and sex affect these prosodic features.

Effect of situation

Towards infants, *f0* mean and range followed the hypothesized pattern, with *f0* mean being higher during the attention getting and *f0* range being wider during the language tutoring situation. This supports the notion that *f0* mean plays a crucial role in controlling and directing infants' attention towards the speaker, while *f0* range contributes significantly to language acquisition (e.g. Trainor & Desjardins, 2002). Conversely, in adult-directed speech, we observed an opposite trend, with speakers using a lower *f0* mean during attention getting compared to the language tutoring situation. This suggests that, with other adults, speakers could use engaging linguistic content rather than relying solely on intense acoustic prosody to capture and maintain their spouses' attention. Interestingly, however, dog-directed *f0* mean showed no difference between the two free speech situations (i.e. attention getting vs. language tutoring). This suggests that speakers did not expect their dogs to form quick object–label associations easily (e.g. Fugazza et al., 2021) and therefore maintained their high pitch to facilitate their canine partner's attention during the language tutoring situation (e.g. Jeannin et al., 2017).

The analysis of the fixed sentences, which contained a nursery rhyme, revealed that speakers utilized the most exaggerated prosody, characterized by higher *f0* mean and wider *f0* range, across all partners (i.e. infants, spouses, and dogs). This finding aligns with our second hypothesis and suggests that the infant-directed nature of the nursery rhyme strongly influenced speech prosody, resulting in a typical rhythmic and melodic speech style with exaggerated acoustics regardless of the partner (e.g. Falk & Audibert, 2021). These results underscore the significance of speech content and its relevance as a factor in the infant-directed nature of a given situation for future comparative prosody research.

Effect of the partners' linguistic competence

In line with our hypotheses, speakers adjusted their speech prosody to their partner's needs and capacities. Specifically, they used a higher and wider ranged *f0* in general, when talking to their infants and dogs compared to when speaking to their spouses. When speakers were attempting to form object–label associations with their infants, they utilized longer utterances (i.e. call length), and female speakers also employed a wider pitch range (i.e. *f0* range). Contrary to our predictions, speakers also used a higher overall pitch (i.e. *f0* mean) when addressing infants while uttering the object label. High pitch might impede word segmentation while also having the potential to capture and maintain infants' attention (Trainor & Desjardins, 2002). It is possible that speakers had to employ more attention-getting cues when uttering the label because infants focused less on the target object, particularly when a non-target object was presented simultaneously. Further analysis of the partner's looking behaviour and attentional states is needed to explore this possibility. When uttering the object label to adults (i.e. their spouses), as expected, speakers used lower mean pitch and smaller pitch range; however, they also employed

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

longer utterances. Previous studies have shown that hyperarticulated vowels and longer utterances are also used towards adults if they are linguistic foreigners (e.g. Uther et al., 2007). Object labels in the present study were artificial words that might resemble foreign phrases, potentially prompting longer utterances from the speakers. Lastly, and in line with our hypotheses, speakers used higher pitch, narrower pitch range, and shorter utterances when uttering the object label to their dogs. These results further support the notion that people tend to adopt a speech style with their dogs aimed at maintaining canine attention, but without the use of language learning aids and likely without word tutoring intentions (e.g. Burnham et al., 2002; Xu et al., 2013; Gergely et al., 2017).

Recently, it has been demonstrated that speakers of the present study express similarly intense happy emotions and emotional valence when interacting with their infants and spouses, while exhibiting less intense and less positive emotions when communicating with their dogs (Koós-Hutás et al., 2024). If the pitch-related features of their speech were to follow this pattern, one could conclude that acoustics are "by-product" of their happy emotions, as previously suggested (Trainor et al., 2000). Our results, however, did not support this notion. Instead, we found that speakers used a higher and more variable pitch when addressing their dogs (and infants) compared to their spouses. This suggests that at least in dog- and adult-directed prosody, the facial and acoustic modalities of prosody exhibit different patterns. These results also suggest that pitch characteristics are not only "by-products" of a more emotional speech style, but also they are functional modifications and are probably adjusted to the partners' emotional needs and cognitive capacities (Trainor et al., 2000; Koós-Hutás et al., 2024).

Effect of the speakers' sex

In line with previous studies, we found more similarities than differences in the acoustic prosody of female and male speakers towards their infants, spouses, and dogs (e.g. Niwano & Sugai, 2003; Gergely et al., 2017). Across situations, both sexes used their *f0* mean and range similarly when speaking to the same type of partner (i.e. infant, spouse, or dog). Moreover, there were no discernible differences between the sexes in the analysis of object labels. In line with prior studies and our hypothesis, the only consistent difference between the two sexes was found in their pitch range: female speakers generally employed a wider *f0* range than male speakers across all partners and situations (e.g. Fernald et al., 1989; Gergely et al., 2017). We also identified minor differences in the *f0* mean of female and male speakers, contrary to our prior expectations. Male speakers, for instance, exhibited a higher *f0* mean when addressing their dogs compared to their infants during the language tutoring situation, while female speakers did not differentiate between partners in terms of f0 mean. Prior research has demonstrated that during tasks involving easy problem-solving, which includes praise, speakers tend to use higher pitch when talking to dogs than to infants (Gergely et al., 2017). It is possible that male speakers praised their dogs more than their infants during the object–label association task or that they required more attention-getting cues to maintain the dog's focus in this setting. Future investigations are needed to test these hypotheses. Moreover, during the fixed sentences situation, female speakers employed a higher mean pitch in their infant-directed speech compared to their dog-directed speech, while male speakers maintained a similar mean pitch when addressing dogs and infants in this scenario. There is evidence that women engage in more frequent singing and rhyming activities with their infants than men, potentially contributing to this discrepancy in the results (e.g. Yan et al., 2021).

https://doi.org/10.1017/S0305000924000552 Published online by Cambridge University Press

Conclusions

The present study supports the well-known phenomenon of more intense acoustic prosodic speech when talking to infants and dogs is still observable when compared to spouse-directed speech. In a comparative framework, we provided further evidence that mean pitch has an important attention-getting function, while pitch range might facilitate language acquisition. Our results suggest that infant-, spouse-, and dog-directed speech prosody conveys more than just positive emotional attitudes; it has the potential to serve specific functions such as capturing attention and aiding language acquisition according to the partners' needs and capacities. Heightened and more variable pitch was found when speakers were reciting a nursery rhyme to both the infant and the dog as well as to their spouses. This finding may indicate that the infant-directed content and context of the speech could have a greater influence on the acoustic prosody than the type of partner. We also found that major patterns of pitch and utterance length modifications are presented similarly in female and male speakers, but female speakers tend to use a wider pitch range in general. In summary, these results highlight the importance of studying the context, content, and addressee-specific features of prosody in a comparative framework to better understand its exact functions and roles.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.17632/z868c5v5yy.1.

Acknowledgements. This study was supported by the Hungarian Scientific Research Fund (NKFIH grant no. FK142968), Hungarian Brain Research Program (HBRP) 3.0 NAP, János Bolyai Research Scholarship (BO/751/20 and BO/00361/24) of the Hungarian Academy of Sciences, and European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (950159). We are grateful to the participating families and to Anna Dallos and Mandula Koós-Hutás for their help in data acquisition.

Competing interest. The authors declare that there are no competing interests.

References

Bates, D., **Mächler, M.**, **Bolker, B. M.**, & **Walker, S. C.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & **Weenink, D.** (2021). *PRAAT. Institute of Phonetic Sciences University of Amsterdam, The Netherlands. Freeware, electronically available.*

Bombar, M. L., & **Littig, L. W.** (1996). Babytalk as a communication of intimate attachment: An initial study in adult romances and friendships. *Personal Relationships*, **3**(2), 137–158. https://doi.org/10.1111/j.1475-6811.1996.tb00108.x

Burnham, D., **Kitamura, C.**, & **Vollmer-Conna, U.** (2002). What's new, pussycat? On talking to babies and animals. *Science (New York, N.Y.)*, **296**(5572), 1435. https://doi.org/10.1126/science.1069587

Chong, S. C. F., **Werker, J. F.**, **Russell, J.**, & **Carroll, J.** (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, **12**, 211–232. https://doi.org/10.1002/icd.286

Cooper, R. P., & **Aslin, R. N.** (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, *Placeholder Text* **65**(6), 1663–1677. https://doi.org/10.1111/j.1467-8624.1994.tb00841.x

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, **27**, 491–541.

Falk, S., & **Audibert, N.** (2021). Acoustic signatures of communicative dimensions in codified mother-infant interactions. *The Journal of the Acoustical Society of America*, **150**(6), 4429–4437. https://doi.org/10.1121/10.0008977

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. *Language and Linguistics Compass*, **16**(1). https://doi.org/10.1111/lnc3.12448

Fernald, A., & **Kuhl, P.** (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., **Taeschner, T.**, **Dunn, J.**, **Papousek, M.**, **de Boysson-Bardies, B.**, & **Fukui, I.** (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, **16**(3), 477–501. https://doi.org/10.1017/S0305000900010679

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Editions de la Maison des Sciences de l'Homme; Cambridge University Press.

Fugazza, C., **Dror, S.**, **Sommese, A.**, **Temesi, A.**, & **Miklósi, Á.** (2021). Word learning dogs (Canis familiaris) provide an animal model for studying exceptional performance. *Scientific Reports*, **11**(1), 1–9. https://doi.org/10.1038/s41598-021-93581-2

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** (1). https://doi.org/10.1038/s41598-017-13883-2

Gergely, A., **Koós-Hutás, É.**, **Filep, L. A.**, **Kis, A.**, & **Topál, J.** (2023). Six facial prosodic expressions caregivers similarly display to infants and dogs. *Scientific Reports* **13** (1), 929. https://doi.org/10.1038/s41598-022-26981-7.

Gogate, L. J., **Bolzani, L. H.**, & **Betancourt, E. A.** (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, **9**(3), 259–288. https://doi.org/10.1207/s15327078in0903_1

Golinkoff, R. M., **Can, D. D.**, **Soderstrom, M.**, & **Hirsh-Pasek, K.** (2015). (Baby)talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, **24**(5), 339–344. https://doi.org/10.1177/0963721415595345

Hirsh-Pasek, K., & **Treiman, R.** (1982). Doggerel: Motherese in a new context. *Journal of Child Language*, **9**(1), 229–237. https://doi.org/10.1017/S0305000900003731

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**(1), 1–9. https://doi.org/10.1038/s41598-017-04671-z

Jürgens, R., **Hammerschmidt, K.**, & **Fischer, J.** (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, **2**, 1–11. https://doi.org/10.3389/fpsyg.2011.00180

Koós-Hutás, É., **Kovács, B. A.**, **Topál, J.**, & **Gergely, A.** (2024). The face behind the caring voice: A comparative study on facial prosodic features of dog-, infant- and adult-directed communication. *Applied Animal Behaviour Science*, **272**, 106203. https://doi.org/10.1016/j.applanim.2024.106203

Lenth, R. V. (2023). *emmeans: Estimated marginal means, aka least-squares means*. R Package Version 1.8.6. https://cran.r-project.org/package=emmeans

Niwano, K., & **Sugai, K.** (2003). Pitch characteristics of speech during mother-infant and father-infant vocal interaction. *The Japanese Journal of Special Education*, **40**(6), 663–674.

Pinheiro, J. C., & **Bates, D. M.** (2000). *Mixed-effects models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Saint-Georges, C., **Chetouani, M.**, **Cassel, R.**, **Apicella, F.**, **Mahdhaoui, A.**, **Muratori, F.**, **Laznik, M. C.**, & **Cohen, D.** (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *PLoS ONE*, **8**(10), 1–17. https://doi.org/10.1371/journal.pone.0078103

Shute, B., & **Wheldall, K.** (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology*, **19**(2), 221–233. https://doi.org/10.1080/0144341990190208

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, **27**(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Thiessen, E. D., **Hill, E. A.**, & **Saffran, J. R.** (2005). Infant-directed speech facilitates word segmentation. *Infancy*, **7**(1), 53–71.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, **85**(5), 1699–1707. https://doi.org/10.1121/1.397959

Trainor, L. J., **Austin, C. M.**, & **Desjardins, N.** (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, **11**(3), 188–195. https://doi.org/10.1111/1467-9280.00240

Trainor, L. J., & **Desjardins, R. N.** (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, **9**(2), 335–340.

Uther, M., **Knoll, M. A.**, & **Burnham, D.** (2007). Do you speak E-NG-L-I-SH ? A comparison of foreigner- and infant-directed speech. *Speech Communication*, **49**, 2–7. https://doi.org/10.1016/j.specom.2006.10.003

Venables, W. N., & **Ripley, B. D.** (2002). *Modern applied statistics with S* (4th ed.). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Warren-Leubecker, A., & **Bohannon, J. N.** (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, **55**(4), 1379–1385.

Weirich, M., & **Simpson, A.** (2019). Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech. *Journal of Speech, Language, and Hearing Research*, **62**(11), 4001–4014.

Woodward, A. L., **Markman, E. M.**, & **Fitzsimmons, C. M.** (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, **30**(4), 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Xu, N., **Burnham, D.**, **Kitamura, C.**, & **Vollmer-Conna, U.** (2013). Vowel hyperarticulation in parrot-, dog- and infant-directed speech. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals*, **26**(3), 373–380. https://doi.org/10.2752/175303713X13697429463592

Yan, R., **Jessani, G.**, **Spelke, E. S.**, **De Villiers, P.**, **De Villiers, J.**, & **Mehr, S. A.** (2021). Across demographics and recent history, most parents sing to their infants and toddlers daily. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**(1840), 1–9. https://doi.org/10.1098/rstb.2021.0089

Cite this article: Koós-Hutás, É., Afrin, S., Kovács, A.B., Faragó, T., Filep, L.A., Topál, J., & Gergely, A. (2025). More than just a happy talk? Evidence for functional pitch and utterance length modifications in infant-, spouse-, and dog-directed communication. *Journal of Child Language* 1–17, https://doi.org/10.1017/S0305000924000552