Acoustic characteristics of dog-directed speech: The role of individual empathy level and personality traits

Jasmin Devi Nuscheler, Nathalie Elsässer, Angelika Braun, Katharina Zahner-Ritter

University of Trier, Department of Phonetics, Germany
jd.nuscheler@web.de, {s2naelsa | brauna | k.zahner-ritter}@uni-trier.de

ABSTRACT

We typically adjust our speech when addressing children or pets. So far, we know very little about the (speaker-specific) factors that explain the degree of such speech style modifications. Here, we compare the acoustic characteristics of speech to a dog (a Mini Maltese, dog-directed speech, DDS) vs. speech to an adult (adult-directed speech, ADS) in German. We test whether the degree of modification between the two speech styles (DDS vs. ADS) relates to a speaker's empathy capacity and personality traits. Results corroborate previous findings on a higher f0 in DDS. Additionally, vowels were fronted and lowered (hyperarticulation), and the postalveolar fricative /ʃ/ tended to be retracted in DDS. The degree of modification between DDS and ADS differed considerably between speakers, but speakers' personality profiles were only weakly correlated with the extent of individual speech style modification – except for spectral cues and conscientiousness, and a tendency for mean f0 and empathy.

Keywords: dog-directed speech, acoustic features, individual variation, empathy, personality

1. INTRODUCTION

A number of studies have shown that speakers phonetically tailor their speech to their interlocutors [1], e.g., when talking to children [2], foreigners [3], or dogs [4]. We know, however, far too little about the speaker-specific factors that determine the degree to which speakers modify their speech. The present study contributes to this new line of research (cf. [4, 5]), focusing on dog-directed speech by non-dog owners in German.

Dog-directed speech (DDS) differs acoustically from adult-directed speech (ADS) [2, 6]: Several studies have reported a higher mean f0 in DDS compared to ADS [6]. Findings on f0 range are inconclusive, with some studies reporting a wider f0 range in DDS than ADS [7], while others do not [4]. Beyond f0 adjustments, [7] and [8] have also demonstrated hyperarticulated vowels in DDS as compared to ADS, i.e., an extension of the vowel space measured via the first (F1) and second formant frequencies (F2), but see [6]. Using higher mean f0 as well as hyperarticulated vowels has been associated with the expression of emotions [8], attracting attention [9], or with the idea of appearing smaller and thus less threatening [2, 10]. Indeed, a dog's size seems to affect speakers' f0, with small dogs being addressed in a higher-pitched voice than big ones [11]. These acoustic differences between ADS and DDS have so far only been studied in dog-owners.

While the acoustic characteristics of different speech styles are quite well understood, we still know very little about the factors that account for inter-speaker variability in the degree of speech style modification [4, 5]. Speakers' personality traits might affect the degree to which they modify their speech: In particular, personality traits have been shown to be relevant with respect to the degree of phonetic accommodation, with openness and neuroticism correlating with how far speakers converge to their interlocutor [12]. For DDS, [4] recently showed that the personality dimension *openness* (positively) predicted the f0 range of female (but not male) dog-owners in DDS. Similarly, a speaker's empathy level might influence acoustic adjustments in DDS. Humans empathise with animals [13], and individual empathy levels affect the way we talk and interpret language [14, 15], for instance with respect to the use of prosodic cues to question intonation or ambiguity resolution [15].

The present study sets out to systematically test whether acoustic modifications between German ADS and DDS also pertain to non-dog owners (**RQ1**). If non-dog owners modify their speech in a similar way to dog owners, we expect higher and more variable mean f0, and more peripheral vowel quality in DDS as compared to ADS.

The second research question (**RQ2**) is more explorative in nature and investigates whether speaker-specific personality traits and a speaker's individual empathy level relate to larger acoustic differences between speech to an adult vs. a dog (in the sense that higher levels of empathy are associated with larger speech style modifications between ADS and DDS etc.).

2. METHODS

Non-dog owners interacted with an adult (ADS) and a 3,5-year-old mini-Maltese, called *Jenna* (DDS, Fig. 1). We used the same dog to control for effects of size [11]. She has a calm and reserved temper and is generally judged as cute. The study was approved by the Ethics Committee of the University of Trier (EK Nr. 35/2022); all participants gave written consent.

2.1. Participants

Twelve German non-dog owners (six male, six female; mean age: 24.3 years, SD = 2.1) participated for a small remuneration. To avoid effects of familiarity, they did not have children, nor did they know the experimenter or the dog. They all showed high affection to dogs, as revealed by a questionnaire (mean = 0.82 on a scale between 0-1, SD = 0.12).

2.2. Materials

Three German nouns were selected to elicit the three cardinal vowels /iː/, /aː/, /uː/ in the two scenarios (see [2, 6, 7]): *Biber* [ˈbiː.bɐ] 'beaver', *Schal* [ʃaːl] 'scarf', and *Schuh* 'shoe' [ʃuː]. Targets were selected such that they were picturable and easy to use as toys. For the picture description task (ADS), 15 (licence-free) pictures were selected (5 pictures for each target). The objects were shown in different contexts (e.g., beaver swimming in a pond or eating some fruit) to increase variation. For the play situation (DDS), we used real objects as toys, i.e., two cuddly toy beavers, two scarfs, and a pair of black sport shoes, see Fig. 1.

2.3. Procedure

Speech recordings. Participants were recoded when interacting with an adult (female, 23 years) and a dog (3,5-year-old mini-Maltese, fixed order) in a quiet room at the University of Trier using a DPA-headset microphone (4088 DCA) connected to a portable recorder (Zoom H4n Pro; 44.1 kHz, 16 bit), which they carried in a little waist bag to be able to move around freely. ADS was elicited via a picture-description task [5] in which participants described 15 pictures to the experimenter (pseudo-randomized such that the same targets did not occur right after each other). In DDS, participants used toys to engage the dog in a shared play, see Fig. 1. Recordings (both phases) lasted about 10 minutes.

Questionnaires. The Big Five Personality Test (B5T, 5-point Likert Scales) [16] was used to assess the five personality dimensions: extraversion (8 items), neuroticism (8 items), conscientiousness (9 items), openness (10 items), and agreeableness (10 items), totalling 45 items.

Participants' empathy capacity was assessed via the *Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie* (SPF, 5-point Likert Scales) [17], which has been used successfully to predict empathy levels in different contexts [18]. The questionnaire includes four subgroups (perspective taking, fantasy, empathic concern and personal distress), but only the first three (16 items in total) are summed up to an individual empathy score [19]. Questionnaires were filled in via *SoSci Survey* [20] after completion of the recordings (15-25 minutes).

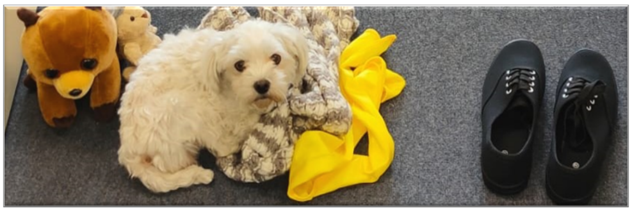

Figure 1: Study-setup in DDS speech style.

2.4. Data processing and statistical analysis

Recordings were manually annotated in Praat [21]: On Tier 1, intonational phrases were segmented, and on Tier 2, the stressed vowels in the three targets were labelled. We additionally labelled the post-alveolar fricative /ʃ/ in *Schuh* and *Schal*, based on an auditory impression of a retracted fricative in DDS.

We excluded phrases in which participants talked in ADS while the dog was present (N = 8), as well as whispered targets (N = 1 in ADS; N = 20 in DDS), see Table 1 for an overview of the dataset.

	ADS	DDS
No. of utterances	1614	1452
No. of target *Biber*	121	79
No. of target *Schal*	103	41
No. of target *Schuh*	221	75

Table 1: Number of utterances and targets per speech style.

Acoustic information was extracted via a Praat script: mean f0 and SD f0 on the phrase level, and the first two formants in the middle of the stressed vowel, as well as the centre of gravity (COG) and its dispersion (SD COG) in the middle of /ʃ/. For formant extraction, we used the linear predictive coding algorithm ([5], t-step = 0.01s, nr of formants = 5, wind. length = 0.025 s, pre-emphasis from = 50 Hz). For male speakers, the maximum formant value was set to 5000 Hz; for females, to 5500 Hz (except for /uː/, where the number of formants was set to 2 and the max frequencies to 1200/1500 Hz). Formant values were manually corrected if necessary. For fricatives, we used a stop Hann band filter to remove f0 before extracting a spectral slice from the middle of the fricative ([22], wind. length: 0.005 s; max freq: 12000 Hz; t-step: 2 ms; freq. step: 20 Hz, wind. shape: Hamming raised sine-squared).

Statistical analyses were run in R(Studio) [23]. To assess the acoustic differences between ADS and DDS (**RQ1**), we used linear mixed regression models. *Speech style* (ADS vs. DDS) was entered as fixed factor in all models. The model for formants additionally contained the factor *vowel*. Random slopes (subject, items) were retained if they improved the fit of the model [24, 25].

To test for a relationship between the personality traits and the degree of modification between the speech styles (**RQ2**), correlation analyses were conducted. To this end, so called difference scores were calculated for each subject for each individual acoustic parameter (e.g., mean f0 in DDS minus mean f0 in ADS). We then correlated these difference scores with the personality trait scores (the empathy level and the five personality dimensions).

3. RESULTS

3.1. Acoustic differences between ADS and DDS

F0. There was an effect of speech style on speakers' mean f0 ($\chi^2(1) = 56.2$, $p < 0.0001$), with DDS being higher than ADS for both female and male speakers (interaction between gender*speech style: $p = 0.14$; males had a numerically larger difference between speech styles). All speakers produced DDS with higher f0, even though the difference between speech style showed considerable individual variation, see Fig. 2. F0 was significantly more variable in DDS than in ADS for male speakers ($ß = 1.39$, $SE = 0.38$, $t = 3.63$, $p < 0.01$), but not for females ($p = 0.1$); the extent of modification again varied across speakers.

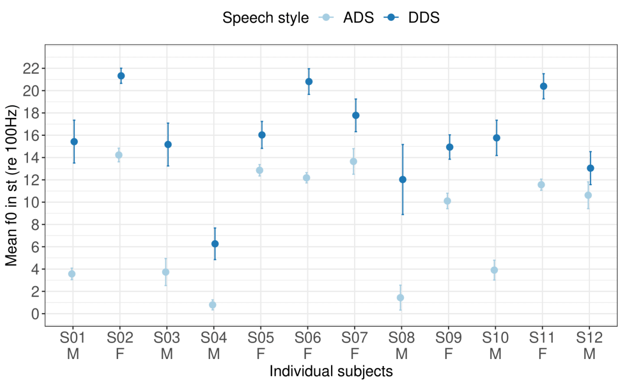

Figure 2: Mean f0 in semitones per speech style.

Vowel quality. Figure 3 shows the averaged vowel space spanned by F1 (vertical dimension, higher values indicate lowered vowel) and F2 (horizontal dimension, higher values indicate fronted vowel). There was no interaction between speech style and gender ($p = 0.23$) or vowel and speech style ($p = 0.13$). However, there was an effect of speech style on the first formant ($\chi^2(1) = 4.6$, $p < 0.05$), which was carried by /u:/ ($p = 0.03$) and /i:/ ($p = 0.06$), and absent in /a:/ ($p = 0.69$). For the second formant, there was an interaction between speech style and vowel ($\chi^2(2) = 22{,}5$, $p < 0.0001$) and no interaction between speech style and gender ($p = 0.11$): All vowels showed a higher second formant, but the difference was smallest for /u:/. However, it must be taken into account that F2-values for /u:/ are relatively small anyway. The distinction for both F1 and F2 across speech styles varied considerably across speakers.

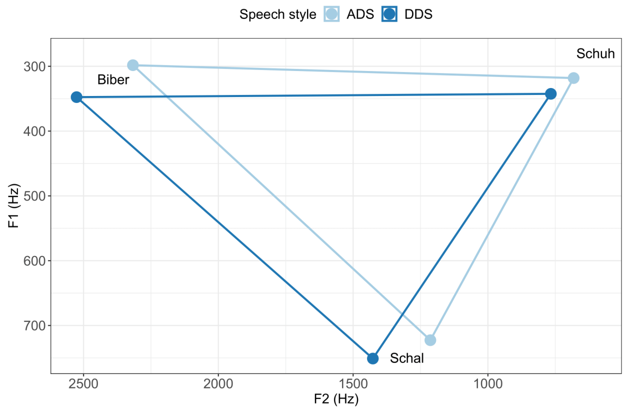

Figure 3: Vowel spaces in both speech styles.

Fricative. An exploratory analysis of the postalveolar fricative /ʃ/ revealed a tendency for the dispersion (SD COG) to be larger in DDS than ADS ($ß = 141.24$, $SE = 69.3$, $t = 2.0$, $p = 0.07$), possibly suggesting a retraction of the fricative. Gender and speech style did not interact ($p = 0.26$); again, participants varied in the extent of modification.

3.2. Relation between personality traits and acoustic modification between DDS and ADS

Figure 4 shows the individual personality profiles (B5T and Empathy) in proportions (0 = lowest, 1 = highest). Reliability of the scales was tested via Cronbach's α, with the empathy subscales ranging between 0.57 and 0.79, and the five dimensions of the B5T between 0.77 and 0.92. Speakers not only differed in the extent of speech style modification (3.1), but also showed considerable variation in their personality traits (see Fig. 4). We now turn to the question of whether the extent of individual differences between DDS and ADS relates to a speaker's personality traits (RQ2).

The correlations between the acoustic changes according to speech style and the personality profiles generally fall short of significance. There is one strong correlation between the dispersion (SD COG) in the postalveolar fricative and the B5T personality dimension *conscientiousness* (r = 0.57, p = 0.05): The more conscientious a participant, the bigger the difference in dispersion between the fricative in DDS and ADS, i.e., the modification of speech style.

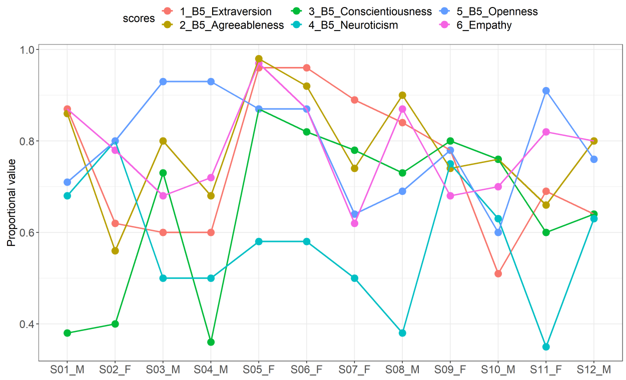

Figure 4: Personality profiles per subject.

On a descriptive level, however, we see some tendencies for relations between personality traits and the extent of speech style modification, e.g., for empathy and the difference in mean f0 (see Fig. 5). Figure 5 shows that for female speakers, higher empathy scores go along with a larger difference between DDS and ADS in mean f0 (red dots; except for one speaker with a high empathy score and a small extent of modification). Post-hoc analyses revealed that larger differences between DDS and ADS in females were due to an increase in f0 in DDS (not ADS) with increasing empathy levels.

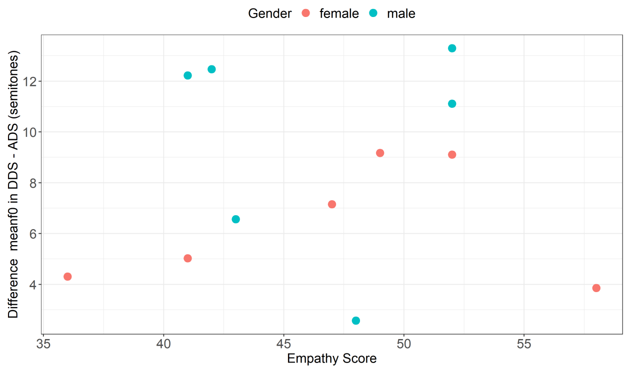

Figure 5: Relation between difference in mean f0 (DDS-ADS) and empathy score.

4. GENERAL DISCUSSION

This study investigated acoustic differences between German DDS and ADS in non-dog owners (**RQ1**), considering the potential influence of personality traits on speech style modification (**RQ2**). Regarding **RQ1**, our findings corroborate previous findings on dog-owners with respect to a higher mean f0 [6]. The participants' general affection to dogs as well as the small size of the dog [11] might have led to this robust difference in all speakers. Results on the variability of f0 were less consistent, showing a larger SD in f0 for male speakers only (see also [4, 7]). Previous studies (on infant-directed speech, IDS) have suggested that an increase in vowel space could be due to increased smiling and hence an overall increase in F2 [26, 27]. Our study, however, revealed differences for both F1 and F2 (lowering and fronting) and is hence most compatible with the interpretation of speakers trying to sound small and non-threatening (shortened vocal tract through raised larynx, cf. [2]), signalled by higher f0 and higher formants [5]. The finding on the spectral distribution in the postalveolar fricative /ʃ/ (tendency towards a larger dispersion of the COG in DDS than ADS), which suggests a backward movement in the place of articulation (i.e., towards the palatal fricative /ç/) [22]), might also be interpreted along these lines: The retraction in place of articulation might reduce the hissing sound in sibilants and hence be an attempt to sound less threatening. This is different from IDS, where sibilants have been shown to be enhanced compared to ADS [28], possibly because IDS serves additional purposes compared to DDS (cf. sound category acquisition [29]). Such interpretations need to be made with care, and future work needs to corroborate this tendency on consonantal differences between DDS and ASD, particularly with regard to the movement of the lips during the interaction with dogs (e.g., smiling). Taken together, our results clearly indicate that non-dog owners (similar to dog-owners) adapt their speech when interacting with dogs (**RQ1**).

The second aim of our study was to test whether speakers' personality traits (B5T dimensions and empathy), factors that have recently been shown to affect phonetic convergence [30] or prosody interpretation [14, 15], account for interspeaker variability in the extent to which speakers adapt their speech in DDS as compared to ADS (**RQ2**). However, speakers' personality traits correlated only weakly with the acoustic modifications, except for conscientiousness and SD COG of /ʃ/. Contrary to [4] we find no relation between openness and f0 range in DDS. Our findings reveal a tendency for higher empathy levels to go hand in hand with larger modifications between DDS and ADS in this group (Fig. 5). Given these tendencies, it is conceivable that relations between personality traits and acoustic cues would be more evident in larger datasets.

To conclude, the results of our study corroborate acoustic differences between speech registers, and at the same time underline the complexity of the relationship between speaker-specific modulations and personality that yet needs to be resolved.

5. ACKNOWLEDGMENTS

The study was financially supported by the *Student Fund* of the Department of Phonetics at the University of Trier.

6. REFERENCES

[1] B. Lindblom, "Explaining Phonetic Variation: A Sketch of the H&H Theory," *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Eds., Dordrecht: Springer, 1990, pp. 403–439.
[2] M. Kalashnikova, C. Carignan, and D. Burnham, "The origins of babytalk: smiling, teaching or social convergence?," *Royal Society Open Science*, vol. 4, no. 8, p. 170306, 2017.
[3] M. Uther, M. A. Knoll, and D. Burnham, "Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech," *Speech Communication*, vol. 49, no. 1, pp. 2–7, 2007.
[4] R. Lesch, K. Kotrschal, I. Schöberl, A. Beetz, J. Solomon, and W. T. Fitch, "Talking to dogs: Companion animal-directed speech in a stress test," *Animals,* vol. 9, no. 7, 2019.
[5] M. Weirich and A. Simpson, "Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech," *JSLHR*, vol. 62, no. 11, pp. 4001–4014, 2019.
[6] D. Burnham, C. Kitamura, and U. Vollmer-Conna, "What's new, pussycat? On talking to babies and animals," *Science*, vol. 296, no. 5572, p. 1435, 2002.
[7] D. Burnham *et al., Are you my little pussy-cat? Acoustic, phonetic and affective qualities of infant- and pet-directed speech*, 1998.
[8] H. J. Kim, M. Diehl, R. Panneton, and C. Moon, *Hyperarticulation in mothers' speech to babies and puppies*, 2006.
[9] A. Fernald and P. K. Kuhl, "Acoustic determinants of infant preference for motherese speech," *Infant Behavior and Development*, vol. 10, no. 3, pp. 279–293, 1987.
[10] J. J. Ohala, "An ethological perspective on common cross-language utilization of F0 of voice," *Phonetica*, vol. 41, no. 1, pp. 1–16, 1984.
[11] Y.-F. Li and P. P. K. Mok, "Does size matter? An preliminary investigation on the effects of physical size on pitch level in pet-directed speech," in *Proceedings of the International Conference on Speech Prosody 2016*, 2016, pp. 1196–1200.
[12] N. Lewandowski and M. Jilka, "Phonetic convergence, language talent, personality and attention," *Front. Commun.*, vol. 4, 2019.
[13] P. Steen, "Selektive Empathie mit Tieren," in *Sprache und Empathie*, K. Jacob, K.-P. Konerding, and W.-A. Liebert, De Gruyter, 2020, pp. 249–284.
[14] N. Esteve-Gibert, A. J. Schafer, B. Hemforth, C. Portes, C. Pozniak, and M. D'Imperio, "Empathy influences how listeners interpret intonation and meaning when words are ambiguous," *Memory & Cognition*, vol. 48, no. 4, pp. 566–580, 2020.
[15] R. Orrico and M. D'Imperio, "Individual empathy levels affect gradual intonation-meaning mapping: The case of biased questions in Salerno Italian," *Laboratory Phonology*, vol. 11, no. 1, 2020.
[16] L. Satow, *SEA - Skala zur Erfassung von Testverfälschung durch positive Selbstdarstellung und sozial erwünschte Antworttendenzen*: ZPID (Leibniz Institute for Psychology Information) – Testarchiv, 2012.
[17] C. Paulus, *Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index*. https://doi.org/10.23668/psycharchives.9249.
[18] S. K. Poppinga, "Empathievermögen und Belastungsempfinden bei Müttern mit physischen Misshandlungserfahrungen in der Kindheit: Selbstauskunft und hirnmorphometrische Korrelate," Charité - Universitätsmedizin Berlin, 2021.
[19] C. Paulus, "Ist die Bildung eines Empathiescores in der deutschen Fassung des IRI sinnvoll?," 2012. http://dx.doi.org/10.22028/D291-23347.
[20] D. J. Leiner, *SoSci Survey. Professionelle Onlinebefragung*. https://www.soscisurvey.de
[21] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. Computer program.
[22] S. Jannedy and M. Weirich, "Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives," *JASA*, vol. 142, no. 1, p. 395, 2017.
[23] R Core Team, *R: A language and environment for statistical computing.*
[24] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.
[25] H. Matuschek, R. Kliegl, S. Vasishth, H. Baayen, and D. Bates, "Balancing Type I error and power in linear mixed models," *Journal of Memory and Language*, vol. 94, pp. 305–315, 2017.
[26] T. Benders, "Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent," *Infant Behav & Development*, vol. 36, no. 4, pp. 847–862, 2013.
[27] E. B. Burnham, E. A. Wieland, M. V. Kondaurova, J. D. McAuley, T. R. Bergeson, and L. C. Dilley, "Phonetic modification of vowel space in storybook speech to infants up to 2 years of age," *JSLHR*, vol. 58, no. 2, pp. 241–253, 2015.
[28] A. Cristià, "Phonetic enhancement of sibilants in infant-directed speech," *JASA*, vol. 128, no. 1, pp. 424–434, 2010.
[29] P. K. Kuhl, "Cracking the speech code: How infants learn language," *Acoust. Sci. & Tech.*, vol. 28, no. 2, pp. 71–83, 2007.
[30] I. Gessinger, E. Raveh, I. Steiner, and B. Möbius, "Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing," *Speech Communica*, vol. 127, pp. 43–63, 2021.