9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Vowel space and f0 characteristics of infant-directed singing and speech

Nicolas Audibert[1], *Simone Falk*[1]

[1]Laboratoire de Phonétique et Phonologie, UMR7018 CNRS/Université Sorbonne-Nouvelle, Paris, France

nicolas.audibert@sorbonne-nouvelle.fr, simone.falk@sorbonne-nouvelle.fr

Abstract

When adults talk to infants, they dramatically change the prosodic and acoustic structure of speech. Recently, new insights have been gained on those changes, especially on the vocalic and temporal structure of speech which are described as being more variable than in adult conversations. In the present contribution, we examine formant and fundamental frequency characteristics of different infant-directed registers, notably infant-directed speech and singing, the latter not being investigated so far. We present data from 14 German-speaking mothers singing a playsong and reading a story to their 6 months old infants, or to the experimenter. Infant- and infant-absent versions of speech and song were compared on the formant characteristics of the primary vowel triangle (/i, a, u/) and on general fundamental frequency changes. Our results show that vowel space did not differ in infant- and infant-absent versions of speech and song. However, vowel dispersion, i.e., formant variability, was higher in both infant-directed song and speech than in infant-absent versions. Consistent with previous findings, f0 was higher in infant- than infant-absent versions of speech and song, with song showing generally higher f0. These results are discussed in light of current approaches to the variability of infant-directed registers, and their attractiveness to infants.

Index Terms: infant-directed registers, speech and singing, infant development, speech acoustics, vocalic space.

1. Introduction

Infants in their first year of life prefer to listen to adult utterances with certain acoustic characteristics such as more variable and higher-pitched f0, longer vowels and a smiling voice timbre [e.g., 1, 2]. These infant-directed (ID) features are used around the world when adults address infants [3]. However, the functions of ID speaking styles are still under discussion. One hypothesis is that ID speech corresponds to highly emotional speech whose purpose is to convey happiness and safety to the infant through its prosodic features [4]. Another hypothesis states that ID speech is a clear and hyperarticulated speech register [5] that could help the infant to build a phoneme inventory of the ambient language [6, 7]. However, this hyperarticulation hypothesis has been controversially discussed in the last years [8]. Although present studies found an enlarged vowel space in ID speech compared to adult-directed (AD) speech when analyzing the primary vowel triangle [6, 7], recent large corpora analyses on ID speech also found reduced non-primary vocalic contrasts and more variable formant instances in ID than in AD speech [8, 9].

These results put into question the clear speech hypothesis of ID registers and incite to examine the role of variability in ID vs. AD registers. Moreover, different ID styles, such as read or spontaneous speech, may show different vowel characteristics according to situational needs. This is equally true for another frequent style at early infant age, namely, infant-directed singing. Although, from an articulatory point of view, singing is clearly different from speaking, ID singing shares many prosodic characteristics with ID speaking. ID singing is higher-pitched and slower than AD singing and appears to have a more loving expressivity [1, 10]. As in speech, infants prefer ID over AD versions of songs [10]. On the other hand, singing generally differs from speaking in that sung vowels usually display more stable f0 contours, and that they are longer and less variable in duration [11, 12]. ID singing also features higher metrical regularity ("beat structure") than ID speech [13]. With respect to vowel formant structure, a corpus study of spontaneous mother-infant interactions [14] showed that ID singing of 5 German mothers displayed an enlarged primary vowel triangle compared to AD speech. However, in this study, no data was available to directly compare ID singing with AD singing or ID singing with ID speech, which makes it difficult to draw conclusions about the vowel space characteristics in ID singing.

The aim of the present study was to examine the vowel space characteristics of ID singing and speaking in comparison to infant-absent (IA) versions. We hypothesized that vowel space in ID versions differs from IA versions, in particular in showing higher formant variability, and potentially, vowel space enlargement.

2. Methods and material

2.1. Corpus

2.1.1. Participants

We recorded 15 mothers, all native speakers of German (age = 31.8 years [*SD* = 3.2])) while they were speaking and singing with their infants. Infants were 6 months old (9 f, 6 m, *M* = 5.8 months [*SD* = 0.9]) at the time of recording. They showed no hearing or other perception or cognitive deficits and were all born on term. Mothers gave informed consent and received a small gift for their participation. All mothers reported to sing regularly with their infants.

2.1.2. Speech material

The material was part of a bigger study. It corresponds to the material described in [15]. In order to measure primary vowels, the words /'bi:ba/ /'ba:bu/ and /'bu:bi/ were chosen. The words were inserted as the names of the protagonists in a German variant of the story ''Three little pigs". They were also inserted in a variant of the German traditional play song "Es tanzt ein Bibabutzemann" (becoming "Es tanzt ein Bibabubibabu"). Mothers were given the texts of the story and

10.21437/SpeechProsody.2018-31

the playsong in advance of the recordings in order to prepare for these new variants. However, all the mothers were familiar with the story and knew the original song.

Mothers were recorded at home. Stories were read from printed text while the playsong was mainly sung without textual support. In two recording conditions, the infant was either present or absent. When the infant was present (ID version), he/she was positioned in close proximity to the mother, mostly sitting or lying on their mother's lap. When the infant was absent (IA version), it was sleeping in another room or being cared for by another person, while the mother read and sang the same material to the experimenter. Recordings were done with an Audio Technica Lavalier Microphone and a Zoom H4-N recorder at 44.100 Hz and a 24-bit sampling rate.

2.2. Acoustic analysis

2.2.1. Segmentation

All the sentences containing the key words were pre-segmented in phones using the automatic segmentation algorithm MAUS [16]. Segmentation errors on the target vowels /i, a, u/ were manually corrected. Disfluencies, slips of the tongues or other errors during reading and singing were excluded from analysis.

2.2.2. Fundamental frequency

Although automated f0 detection procedures in Praat [16] can achieve accurate f0 detection in most cases, ID speech and singing are produced with a highly variable pitch in addition to an overall increased f0 level, making detection errors more likely.

We therefore used a custom procedure to define detection parameters for each speaker*style (song / speech), in order to minimize f0 detection errors. The retained approach was inspired by [18], in which the acceptable range of f0 values was defined relative to quantiles in the distribution of values detected with default parameters. In a first step based on a custom Praat script, f0 detection was performed iteratively, showing the distribution of f0 values and letting the user adjust detection parameters if needed. In a second step, the lowest and highest detected f0 values for each speaker*style were visually inspected in the context of the matching utterance to check for possible remaining octave jumps. f0 values retained in the analysis were extracted at the midpoint of each target vowel.

2.2.3. Formant measurements and vocalic space metrics

F1 and F2 vowel formants were measured at the temporal midpoint of the target vowel (i.e., /a, i, u/, accented and non-accented in the key words /'bi:ba/ /'ba:bu/ and /'bu:bi/), using Praat's implementation of the Burg method with a set of parameters chosen to minimize the number of erroneous values: 5.5 formants with a maximum frequency of 5500 Hz. F1 and F2 values were automatically checked against standard formant values for /a, i, u/ in German and filtered when largely deviating from those values. Overall, 8% of the data were discarded. In addition, one mother was excluded from further analyses as too many vowel occurrences in singing had to be discarded due to measurement errors. Thus, the final data set included the data of 14 mothers and 9,875 occurrences of the target vowels. Given the expected neutralization between accented and unaccented vowels in the singing condition, only 3 vowel classes /a, i, u/ were retained, for both filtering and analysis. All formant values were converted to Bark scale [19] prior to statistical analysis in order to avoid giving an artificially high weight to higher formant values in distance computation.

Since the different dimensions of vocalic variation can hardly be described by a single measure (see for instance [20] or [21]), two separate metrics were computed on Bark-transformed formant values, following [22]:

(1) The distance *DistCentroid* of each vowel to the vowel space centroid, computed as the mean of centroids for /a/, /i/ and /u/ for each speaker*style (song/speech). When computed over a set of vowels, this measure gives an estimation of the overall dispersion of the vocalic space, similarly to the classical area of the vocalic triangle or polygon. In our data, a fairly high correlation (r=.93) is indeed observed between the mean distance to vowel space centroid for each speaker*style and the area of the vocalic triangle.

(2) The distance *V-Dispersion* of each vowel to the centroid of the vowel category (e.g. the centroid of /a/ if the vowel is an [a]), also computed for each speaker*style. This measure gives an estimation of the variability within each vowel category.

3. Results

3.1. Qualitative inspection of vocalic spaces

Vocalic spaces presented on Figure 2 were plotted using the phonR package [23]. Visual inspection of the vowel centroids positions suggests that in both spoken and sung style, the main correlates of the presence of the infant are an expansion on F1 and a compression on F2, although the latter is more clear-cut in the sung condition. On the other hand, within-vowel variability appears larger in both conditions when the infant is present, for each vowel category.

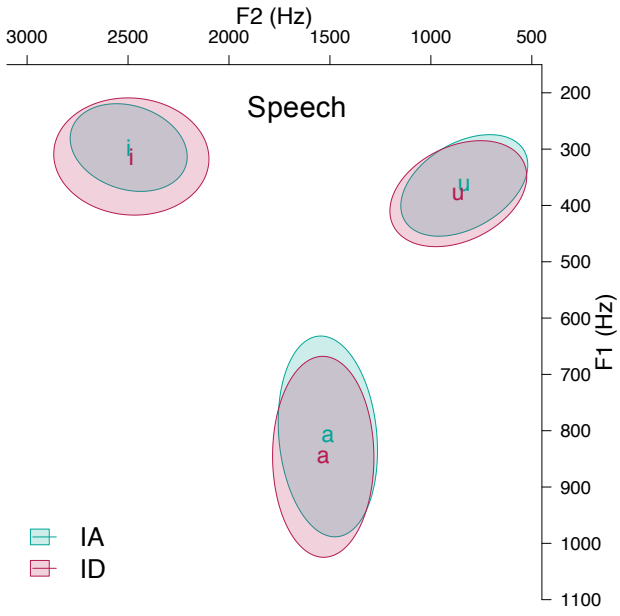

Figure 1: *Vocalic spaces in the F1/F2 plane in spoken style for IA = Infant-absent; ID = Infant-directed versions. Ellipses represent 1 standard deviation in the bivariate space (68% confidence interval).*

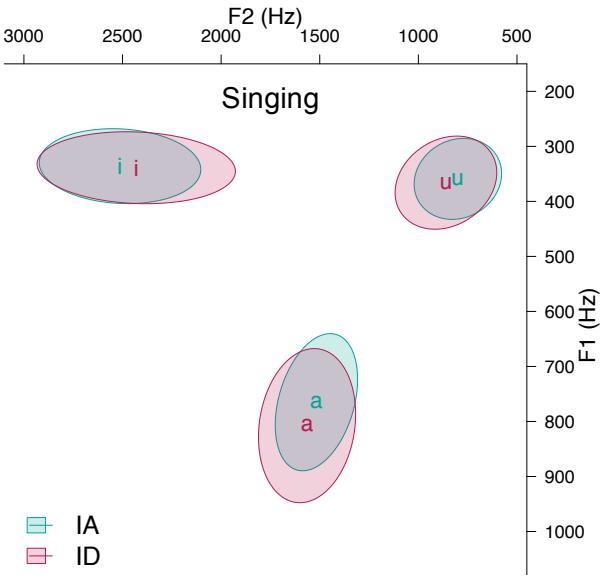

Figure 2: *Vocalic spaces in the F1/F2 plane in singing style for IA = Infant- absent; ID = Infant-directed versions. Ellipses represent 1 standard deviation in the bivariate space (68% confidence interval).*

3.2. Statistical analysis

Figures 3 to 6 illustrate style differences between speaking and singing and version differences when the infant was present or absent in the productions of the 14 mothers, for the variables f0, Duration, DistCentroid and V-Dispersion. In those figures, error bars represent the standard error of the mean.

A linear mixed effects model was fit using R package lme4 [24] for dependent variables f0, Duration, DistCentroid and V-Dispersion, using Style (singing/speaking), Version (IA, ID) and Vowel (/a, i, u/) as fixed effects, and Speaker as random effect. Since the analyzed material is expected to show large variation of fundamental frequency, f0 was included in the model as a random effect for the dependent variables DistCentroid and V-Dispersion. Results presented here focus on the effects of Version and Style.

Given the large number of vowels analyzed, *p*-values estimated using the log-likelihood test are very low for all dependent variables and fixed factors (the highest being p=.005 for the effect of Version on DistCentroid, some values being too low to be accurately estimated), and therefore do not give much useful information. Therefore, effect sizes were estimated by the marginal R^2 associated with fixed effects as in [25]. Computation of marginal R^2 values was performed using the R package r2glmm [26].

As expected, the largest marginal R^2 values are found for effects on f0, with a larger effect of Style (R^2=.210) than the effect of Version (R^2=.024), and for effects on Duration (R^2=.288 for the effect of Style, R^2=.019 for Version). Considering effects on vocalic space, both Style and Version have a much larger effect on V-Dispersion (R^2=.012 for the effect of Style, R^2=.010 for Version) than on DistCentroid (R^2=.001 for the effect of both Style and Version), indicating that those factors affect within-vowel variability rather than the overall vocalic space size. However, Figure 5 suggests that, unlike within-vowel variability which is increased in the presence of the infant in both styles, differences in vocalic space size between versions would be specific to singing.

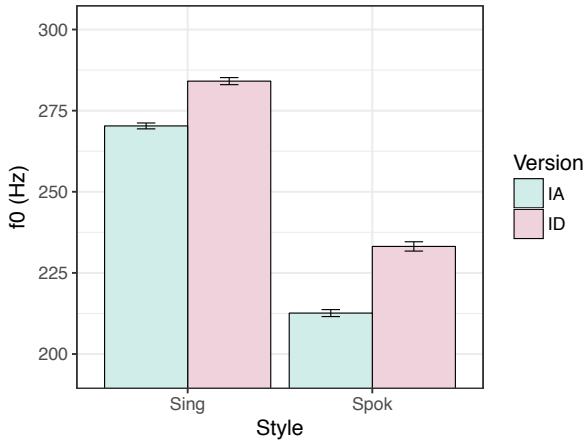

Figure 3: *f0 levels in Hertz in infant-present and –absent versions of singing and speaking, all speakers pooled.*

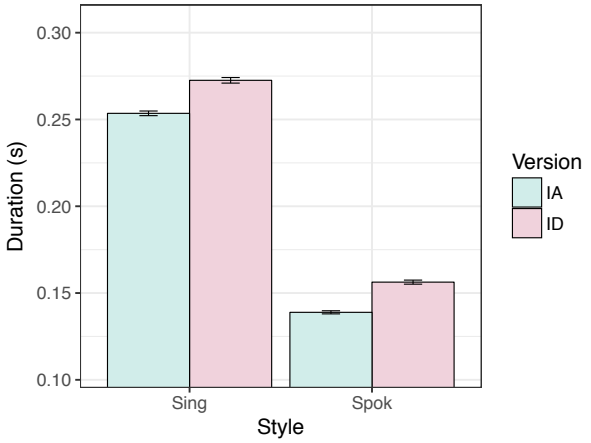

Figure 4: *Mean duration in seconds in infant-present and –absent versions of singing and speaking, all speakers pooled.*

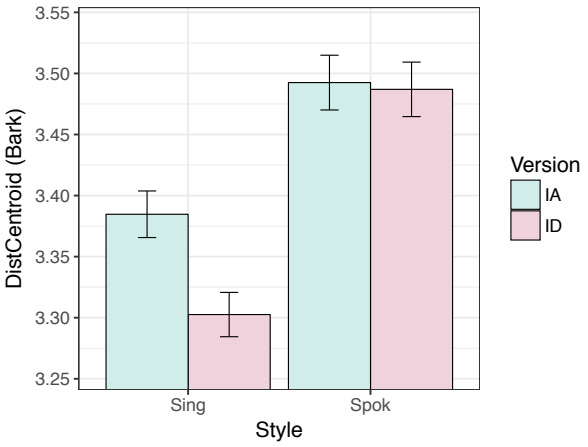

Figure 5: *Vocalic space dispersion (in Bark, estimated by DistCentroid) in infant-present and –absent versions of singing and speaking, all speakers pooled.*

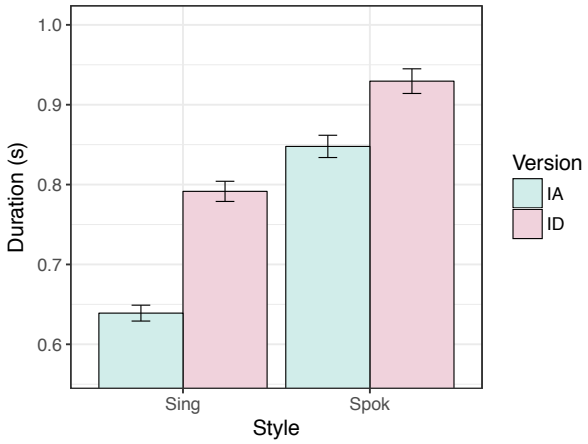

Figure 6: *Within-vowel-category variability (in Bark, estimated by V-Dispersion) in infant-present and – absent versions of singing and speaking, all speakers pooled.*

4. Discussion

We compared ID and IA versions of speaking and singing on their formant and f0 characteristics. Our results showed that differences between ID and IA versions were most visible in overall f0 level, and to a lesser extent in within-vowel-category variability. Both acoustic dimensions were overall higher in ID than IA versions, consistently with previous studies [8, 27]. In contrast, vowel space did not differ much between ID and IA versions in our sample contrary to [6, 7, 8, 28]. One explanation given by McMurray et al. [29] could be that higher formant variance counteracts vowel space expansion in ID registers. In sum, from our analyses, we do not find evidence for supporting a clear account of ID registers as hyperarticulated registers. However, we find evidence for ID registers as acoustically more varied registers.

What is the reason for the pervasiveness of higher variability in ID registers? Similar to formants, temporal and amplitude dynamics as well as pitch are more variable and more contrastive in ID speech and playful singing than in IA renditions [1, 3, 15, 30]. Several groups of researchers [8, 29, 31] propose that changes in segmental characteristics in ID speech (such as vowel variability, expansion or slower tempo) are largely the outcome of suprasegmental phenomena such as emphasis of prosodic boundaries or overall shorter phrases. Therefore, variability in segmental characteristics may be best investigated in relation to the acoustic and prosodic hierarchical structure of speech (see e.g., [15]). An interesting theoretical account of variability is provided by Eaves et al, [32]. Here, learning algorithms evaluating the optimal input for learning phonetic categories in American English showed an advantage for higher formant variability, less hyperarticulation of corner vowels and even hypoarticulation of some categories as found in real ID speech. Thus, higher vowel variability in ID styles may not contradict accounts of higher learnability of ID than IA or AD registers by infants, similar to enhanced phonological learning in infants from higher speaker variability (e.g., [33])

Speaking and singing also showed significant differences. Mean f0 was higher when mothers sang the playsong compared to story reading which may be partly due to the specific melodic structure of the song (i.e., a large pitch range) and the content of the story (e.g., mothers uttered the wolves' statements with a very low f0 level). Previous studies were not conclusive about higher pitch in ID speech or singing [13]. Importantly, within-vowel-category dispersion was higher in speech than in song. This is consistent with the idea that singing compared to speaking provides more stability in the vowel domain, in duration, but also in pitch and formant structure [11, 12]. At the same time, vowel space expansion appeared overall smaller in song than in speech, mainly due to more closed articulation of /a/ in singing. This could be the reflection of a strategy to better control the loudness of /a/ through smaller jaw opening. As all mothers were amateur singers, loudness control through air flow whose intensity has to be fine-tuned relative to pitch height [34] may be more difficult for them as its mastery requires vocal training. Alternatively, singing mothers may have been more smiling compared to speaking (as suggested by [35]), which could also diminish jaw opening and therefore, F1.

As a side note, although we did not test this directly, it should be noted that the difference in within-category vowel variability between IA and ID singing seems to be more important than the difference in variability between IA and ID speaking. Similarly, inspection of the overall vowel space expansion suggests that vowel space is a little more compressed in the F2 plane in ID than IA versions in song, while no such difference was preeminent for IA vs. ID speech. While duration differences were not the focus of the present study, they are consistent with findings in previous studies, with longer durations for infant-directed and sung productions. Since increased durations are expected to give rise to larger vowel spaces, those differences are unlikely to explain solely the observed patterns.

Qualitative inspection of between-speaker differences suggest that, while global trends are shared by most mothers, some of them seem to have different strategies when adapting their productions to the presence or absence of the infant and/or to different styles. These individual differences will be addressed in future work.

In order to better understand the different structures of ID acoustics in ID speech and song, future studies should investigate these differences in more detail and potentially complement them with articulatory data.

5. Acknowledgements

We would like to thank Anne Zorn, Veronika Neumeyer and Elena Maslow for help with data collection and analyses. This work was funded by a MCRI-SSHRC Canada (www.airsplace.ca) research grant and supported by the program "Jeune chercheur" of the University Sorbonne Nouvelle, as well as the Labex EFL program (ANR-10-LABX-0083).

6. References

[1] Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior & Development, 20*, 383–396.

[2] Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development, 10*, 279–293

[3] Fernald, A., Taeschner, T., Dunn, J., Papoušek, 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.

[4] Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*, 365–394

[5] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer Netherlands.

[6] Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science,* 296, 1435.

[7] Kuhl, P, Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. L., Sundberg, U, & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science,* 277, 684–686.

[8] Cristia, A., & Seidl, A., 2014. The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language,* 41, 913–934.

[9] Martin, A., Schatz, T., Versteegh, M., Miyazawa K., Mazuka R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticualtion hypothesis. *Psychological Science,* 26, 341–347.

[10] Trainor, L. J. (1996). Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behavior & Development, 19*, 83–92.

[11] Falk, S., Maslow, E., Thum, G., & Hoole, P. (2016). Temporal variability in sung productions of adolescents who stutter. *Journal of Communication Disorders*, 62, 101–114

[12] Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants prefer infant-directed song over speech. *Child Development*, *88*(4), 1207-1215.

[13] Bergeson, T., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science, 13*, 72-75.

[14] Falk, S. (2007). Speech Clarity in Infant-directed Singing: an Analysis of German Vowels. *Proceedings of the XVI. ICPhS Conference, Saarbrücken, Germany.*

[15] Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-directed speech and song. *Cognition, 163*, 80–86.

[16] Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. *Proceedings of the ICPhS 1999*, 607–610.

[17] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[18] De Looze, C., & Hirst, D. J. (2008). Detecting changes in key and range for the automatic modelling and coding of intonation. *Proceedings of the Speech Prosody 2008 Conference, Campinas, Brazil*, 135-138..

[19] Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97-100.

[20] Ferguson, S. H., Kewley-Port, D. 2007. Talker Differences in Clear and Conversational Speech: Acoustic Characteristics of Vowels. *Journal of Speech, Language, and Hearing Research*, Vol. 50, 1241–1255.

[21] Harmegnies, B., Poch-Olivé. 1992. A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish. *Speech Communication*, 11, 429–437.

[22] Audibert, N., Fougeron, C., Gendrot, C., & Adda-Decker, M. (2015, August). Duration-vs. style-dependent vowel variation: A multiparametric investigation. In *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, Scotland*

[23] McCloy, D. R. (2016). phonR: tools for phoneticians and phonologists. R package version 1.0-7.

[24] Bates, D., Maechler, M., Bolker, B., & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

[25] Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

[26] Jaeger, B. (2017). r2glmm: Computes R Squared for Mixed (Multilevel) Models. R package version 0.1.2.

[27] Nakata, T., & Trehub, S. E. (2004): Infants' responsiveness to maternal speech and singing. In: *Infant Behavior and Development* 27, S. 455–464.

[28] Kalashnikova, M, Carignan, C., & Burnham, D. (2017). The origins of babytalk: smiling, teaching or social convergence? *Royal Society of Open Science,* 4: 170306.

[29] McMurray, B., Kovack-Lesh, K., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129,* 362–378.

[30] Nakata, T., & Trehub, S.E. (2010). Expressive Timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicology: Music, Mind & Brain*, 21, 130-137.

[31] Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, 156, 52–59.

[32] Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, 123, 758-771.

[33] Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12, 339-349.

[34] Sundberg, J. (1987): *The Science of the Singing voice*. Illinois : Northern Illinois University Press.

[35] Trehub, S. E., Plantinga, J., & Russo, F. A. (2016). Maternal vocal interactions with infants: Reciprocal visual influences. *Social Development, 25*.