JASA ARTICLE

Acoustic signatures of communicative dimensions in codified mother-infant interactions

Simone Falk[1,a)] and Nicolas Audibert[2]

[1]Département de Linguistique et de traduction, Université de Montréal, 3150, rue Jean-Brillant, Montreal, Quebec, H3T 1N8, Canada

[2]Laboratoire de Phonétique et Phonologie, UMR7018, CNRS/Université Sorbonne-Nouvelle, 19 Rue des Bernardins, Paris, 75005, France

ABSTRACT:

Nursery rhymes, lullabies, or traditional stories are pieces of oral tradition that constitute an integral part of communication between caregivers and preverbal infants. Caregivers use a distinct acoustic style when singing or narrating to their infants. Unlike spontaneous infant-directed (ID) interactions, codified interactions benefit from highly stable acoustics due to their repetitive character. The aim of the study was to determine whether specific combinations of acoustic traits (i.e., vowel pitch, duration, spectral structure, and their variability) form characteristic "signatures" of different communicative dimensions during codified interactions, such as vocalization type, interactive stimulation, and infant-directedness. Bayesian analysis, applied to over 14 000 vowels from codified live interactions between mothers and their 6-months-old infants, showed that a few acoustic traits prominently characterize arousing vs calm interactions and sung vs spoken interactions. While pitch and duration and their variation played a prominent role in constituting these signatures, more linguistic aspects such as vowel clarity showed small or no effects. Infant-directedness was identifiable in a larger set of acoustic cues than the other dimensions. These findings provide insights into the functions of acoustic variation of ID communication and into the potential role of codified interactions for infants' learning about communicative intent and expressive forms typical of language and music.
 https://doi.org/10.1121/10.0008977

(Received 18 June 2021; revised 11 November 2021; accepted 20 November 2021; published online 21 December 2021)

[Editor: Susanne Fuchs] Pages: 4429–4437

I. INTRODUCTION

Communication with preverbal infants constitutes a challenge for adults. Caregivers have to convey communicative intent and regulate infants' states without having access to symbolic common ground and representations. As a consequence, caregivers of young infants often resort to codified interactions from child-lore and oral tradition, handed down from generation to generation. Sung and spoken games, rhymes, lullabies, and stories are part of this rich cultural repertoire which often accompanies children and parents throughout infancy and early childhood [e.g., Opie and Opie, (1997)]. Codified interactions become quickly ritualized in mothers' and infants' daily routines and are clearly associated with specific communicative situations (Trehub and Gudmundsdottir, 2019). For example, caregivers around the world sing lullabies to infants under distress or before bedtime to soothe and calm them (Mehr *et al.*, 2018; Trehub and Trainor, 1998). Playsongs or rhymes, often accompanied by finger games, tickling or other body movement, arouse infants and enhance reciprocity in mother-infant interactions during playtime (Rock *et al.*, 1999; Vlismas *et al.*, 2013). Many parents in Western societies start narrating stories as early as after birth as part of their daily routines, with potential benefits for children's word learning and later reading capacities (Montag *et al.*, 2015).

Codified interactions are particularly appropriate to foster joint action (Phillips-Silver and Keller, 2012), dyadic reciprocity (Malloch, 1999; van Puyvelde *et al.*, 2010), social learning [e.g., Mehr *et al.* (2016) and Ramírez-Esparza *et al.* (2014)] and regulate infants' emotional state and arousal levels [e.g., Cirelli *et al.* (2020) and Corbeil *et al.* (2016)]. In light of these benefits, it seems surprising that, given the rich literature on infant-directed communication, only a small part of the research focuses on the acoustic traits of codified interactions. Indeed, the ritualized character of codified interactions entails high numbers of repeated performances resulting in highly stable acoustics over time, featuring similar tempo and pitch height, even when performed days apart (Bergeson and Trehub, 2002). Such stable associations between acoustic form and function (situational context) have high potential to provide an acoustic "signature" to infants, that is, combinations of acoustic traits that are more representative of one communicative dimension than another. The presence of such acoustic signatures in codified interactions could be a valuable source of information to guide infants' attending and foster their capacity to predict communicative intent as well as to develop interactive and self-regulatory skills [e.g., Kalashnikova *et al.*

[a)]Also at: International Laboratory for Brain, Music, and Sound Research (BRAMS), Université de Montréal, 90 Vincent D'Indy Ave, Outremont, QC H2V 2S9, Canada. Electronic mail: simone.falk@umontreal.ca

(2017) and Mehr *et al.* (2018)]. Therefore, the aim of the present study was to investigate the presence of such acoustic signatures, by directly comparing three communicative dimensions—interactive stimulation (arousing or calm), vocalization type (speech or song), and infant-directedness [infant-directed (ID) or adult-directed (AD)].

Previous research has investigated these communicative dimensions separately from each other, pointing out pitch, rhythmic, and segmental properties that could indeed form potential characteristic signatures. Interactive stimulation refers to the communicative intent of the interaction, that is, whether a playful and arousing or a calm and soothing performance is intended with the codified interaction. Note that codified interactions are very clearly defined on one pole of the arousing/playful—soothing/calm scale because of the high repetitiveness of the interactions during daily routines. Pitch contour has been reported as a particularly prominent trait of this dimension. In ID speech and song, adults from different cultural and linguistic contexts (i.e., American English, German, French, Russian, Mandarin) frequently use steeply ascending intonation or melodic contours during arousing play situations, while pitch contour composition clearly differs in calm situations (Falk, 2011; Fernald, 1989; Papoušek *et al.*, 1991). In ID singing, lower pitch associated with slower tempo is characteristic of soothing songs and speech, while higher pitch and faster tempo are associated with play situations (Mehr *et al.*, 2018; Trehub and Trainor, 1998). Infants prefer these situation-specific combinations of acoustic traits [e.g., slow and low for lullabies, and fast and high for playsongs, Conrad *et al.* (2011) and Tsang and Conrad (2010)]. Currently, no studies are available that have examined segmental properties of ID communication as a potential marker of this dimension.

As the second communicative dimension, ID acoustics can help identifying the type of vocalization, speech, or song, which caregivers use when addressing infants. Although this distinction may be evident for older children and adults [see Tsang *et al.* (2017)], a distinction between language and musical sounds seems not to be present at birth (Kotilahti *et al.*, 2010) and young infants still need to learn about the differences (Fava *et al.*, 2014). Infants are specifically attracted to the acoustics of ID song compared to ID speech between 5 and 10 months of age (Nakata and Trehub, 2004; Tsang *et al.*, 2017), although at other time points during their first year of life, they have shown similar interest in either type of vocalization (Corbeil *et al.*, 2013; Costa-Giomi and Ilari, 2014). Some distinctive features of ID singing are overall slower tempo, smaller pitch range, and lower variability of pitches within a vowel compared to ID speech [e.g., Tsang *et al.* (2017)]. In addition, ID songs are more rhythmical than ID speech, and pitch height (i.e., mean f0) and tempo vary less in ID singing then in speech across repetitions (Bergeson and Trehub, 2002). Vowel space in Western classical singing has been found to be more compressed than in speech, and vowels as being less variable (Bradley, 2018). ID singing as well has occasionally been reported to show less variable vowels than ID speech (Audibert and Falk, 2018), although further research is needed to reach conclusive results.

Finally, the third communicative dimension concerns the differences in pitch, timbre, tempo, and sound characteristics between infant-directed vs adult-directed communication [e.g., Farran *et al.* (2016), Fernald *et al.* (1989), and Piazza *et al.* (2017)]. Pitch and timbral structure in ID communication differs in a major way from AD communication as caregivers display overall higher pitch, a "smilier" voice and, at least in spontaneous ID speech, expanded pitch contours compared to AD interactions [e.g., Fernald *et al.* (1989), Narayan and McDermott (2016), and Trainor *et al.* (1997)]. Temporal structure in both ID speech and song features more and shorter phrases as well as longer phrase-final syllables and pauses, resulting in overall enhanced temporal hierarchical structure in ID vs AD interactions (Falk and Kello, 2017; Martin *et al.*, 2016). Vowel sounds display different qualitative characteristics in ID vs AD interactions, at least in speech. Many studies, although not unequivocally, have reported more separable focal vowels /a, i, u/ [e.g., Burnham *et al.* (2002) and Kuhl *et al.* (1997)]. In addition, larger within-category variability in ID vs AD utterances throughout the vocalic system has been reported [e.g., Martin *et al.* (2015) and McMurray *et al.* (2013)]. Generally, these acoustic modifications may serve the purpose to attract and maintain infants' attention during interaction. Infants prefer to listen to ID compared to AD utterances, even in the absence of visual or tactile information [e.g., Cooper and Aslin (1990) and Masataka (1999)] and are particularly sensitive to pitch and timbral modifications [e.g., Fernald and Kuhl (1987)].

Based on these findings, pitch height and variability should be among the most prominent traits in the acoustic signatures of infant-directedness, as well as interactive stimulation. Temporal structure and its variability should constitute a prominent marker of vocalization type, that is, differences between speaking and singing, and, maybe to a lesser extent, interactive stimulation. Segmental variation, especially vowel clarity of focal vowels, should most likely change as a function of infant-directedness, while vowel variability could be related to any of the three dimensions.

To investigate these issues, we recorded mothers' performances of codified interactions, a spoken rhyme, a playsong, a lullaby, and a popular story, in the presence (ID) or absence (AD) of their 6-month-old infant in a semi-spontaneous experiment. At this infant age, just before the onset of independent motion as well as before infant's canonical babbling phase/first word production, mothers and infants fully benefit from codified interactions to sustain their communication.

Pitch, duration, and vowel quality as well as their variability were extracted from over 14 000 vowels. Bayesian regression was used to estimate the probability of an effect of infant-directedness (ID vs AD interactions), interactive stimulation (arousing play vs calm listening) and vocalization type (spoken vs sung expressions) on these acoustic features. These estimates were used as an index of the contribution of each acoustic variable to the acoustic signature of each communicative dimension.

II. METHODS

A. Participants

Fifteen mothers (mean age = 31.8 years, SD = 3.2) with infants aged 6 months (9 f, mean age = 5.8 months, SD = 0.9) were recruited in the Munich area, Germany. All of them were native speakers of German having the habit of regularly singing with their infants during the last month (multiple times or at least once per day). They were all familiar with the songs, rhymes, and the content of the story they performed during recordings. Infants were born on term and showed no deficits in hearing, cognitive and motor development. Mothers gave informed consent and received a small gift for their participation. Ethical conduct was in line with the Declaration of Helsinki.

B. Materials and Procedure

Mothers performed two types of spoken (a rhyme, a read story) and two types of sung (a playsong, a lullaby) samples of traditional German child-lore. The samples were very common and representative of early codified interactions in a Western cultural context and chosen for their possibility to induce different interactive styles. Playful, sensorimotor arousing interactions (such as rhythmically bouncing the baby, tickling, moving body parts) eliciting infants' vocal participation were expected to occur when parents performed the spoken rhyme and playsong [e.g., Falk (2009) and Trehub and Trainor (1998)], while story reading, and the lullaby were expected to yield a calm performance without or with significantly fewer motor components and infant vocalizations. Mother's actual ID performances were documented by the experimenter, in form of an observational protocol for each recording session. These protocols were scored for absence (0) or presence (1) of above motor components and, in addition, for presence (0) or absence (1) of infants' vocal participation in the performance (cooing, laughing, vocalizing; apart from crying or fuzziness). Results (displayed in Table I) confirmed that playful motor stimulation and infant co-vocalizations were significantly more frequent during rhyme/playsong performances vs story/lullaby performances (Wilcoxon signed ranks; $Z = -2.89$, $p = 0.004$).

TABLE I. Interactive score rated on a 0-2 scale for each condition, taking into account motor components in mothers' performances and infant co-vocalizations. Mean value are presented with standard deviation in parenthesis.

Interactive stimulation \ Vocalization type	Speech	Song
arousing	rhyme: 1.0 (0.38)	playsong: 1.1 (0.70)
calm	story: 0.7 (0.50)	lullaby: 0.4 (0.51)

To obtain comparable words and vowels (/a, i, u/) across interactions for acoustic evaluation, the names of the protagonists of the sung and spoken child-lore samples were replaced by the names "Biba, Babu, Bubi" (see supplemental material[1]; note that while "Biba" and "Babu" are real nonce names in German, "Bubi" is a word that can appear in the Bavarian dialect as a diminutive form for "small boy"). Mothers were informed in advance of these changes in order to prepare and practice for the recording. The recordings (via an Audio Technica Lavalier Microphone and a Zoom H4-N recorder at 44.100 Hz and a 24-bit sampling rate) took place in two sessions (one infant-present, one infant-absent, counterbalanced across participants) at the mother-infant dyad's home (duration between 0.5 and 1.5 h per session), in the presence of the same female experimenter. In ID versions, the infant was positioned in close proximity to the mother (e.g., sitting or lying on the mother's lap). When the infant was absent (e.g., sleeping or cared for by another person in another room), the mother read and sang the same material to the experimenter (AD version).

C. Acoustic analysis

Longer passages containing disfluencies, slips of the tongue or other errors during speaking and singing were discarded prior to the analyses. All vowels in the target words /bi:ba/, /ba:bu/, and /bu:bi/ were pre-segmented using a semi-automatized procedure, resulting in an initial 17 195 vowels (MAUS) (Schiel, 1999). Please note that, for the sake of simplicity, all vowels in these words are summarized here under the labels /a, i, u/ irrespective of their accentuation and phonological length (in German, the long vowels /a:, i:, u:/, and the length contrast in general, can only occur in accented syllables). All audio files were checked for background noises stemming from the infant, the mother or the technical equipment. After inspection, around 200 vowels were excluded because of noise interference.

All acoustic analyses were performed using PRAAT (Boersma, 2001). For each vowel, vowel duration was extracted from the segmentation. Vowel duration variability was calculated for each vowel as the normalized difference between the duration of the vowel and the mean value per vocalization type (song/speech) and speaker. Pitch height was determined by extracting fundamental frequency (f0) at each midpoint of a target vowel (/a, i, u/), using the autocorrelation-based pitch detection method implemented in PRAAT (Boersma, 1993). In order to minimize detection errors, f0 detection was constrained by adjusting the accepted range of values for each speaker*vocalization type. Following a procedure inspired by De Looze and Hirst (2008), adjusted values were defined based on the distribution of values obtained from a first detection with default settings. Obvious f0 detection errors (octave jumps) were discarded based on the visual inspection of the distribution (~60 vowels). Pitch (f0) variability was computed as the difference in semitones between the f0 value for each vowel and the mean value per vocalization type and speaker.

Vowel quality was measured via the F1 and F2 vowel formants (Burg method, five formants with a maximum frequency of 5500 Hz) at the temporal midpoint of the target vowel and converted to Bark scale. Vowel clarity was expressed as the Euclidean distance of each vowel to the vowel space centroid, assuming that larger distance represents higher vowel clarity [see, e.g., Bradlow *et al.* (1996)]. The vowel space centroid was derived from the mean centroids of all occurrences of /a/, /i/, and /u/ for each speaker*vocalization type (song/speech). Vowel variability within each category was estimated as the distance of each vowel to the centroid of the vowel category (e.g., the centroid of /a/, if the vowel is an accented or unaccented exemplar of /a/), also computed for each speaker*vocalization type. Vowel formant values were automatically checked against standard values for /a, i, u/ in German female speakers (Pätzold and Simpson, 1997) and excluded when largely deviating from those values (see Table II for the thresholds). Thresholds were set to allow for larger variations in vowel realization compared to expected formant values in read speech. Data of one mother had to be discarded due to too many formant measurement errors in sung performances (up to 53% of erroneous values for the infant-directed lullaby), which seem partly linked to the increased difficulty of automated formant estimation with higher f0 as outlined by Maurer (2016). Another 1037 vowels were discarded in the remaining participants due to gross formant detection error. After these corrections, the final number of vowels for statistical analysis was 14.519 (for the final data set, see supplemental materials[1]).

D. Statistical analysis

In recent years, Linear Mixed-effects models (LMM) became one of the most widespread statistical approach to acoustic data analysis (Baayen *et al.*, 2008). However, there are a number of limitations to this family of frequentist analyses. First, fitting LMMs to data with complex random structure often leads to computational issues with models failing to converge. Second, when using analyses in which data interpretation is based on significance values, large datasets give rise to the problem that even small effects tend to be highly significant. In contrast, a Bayesian modeling approach allows one to circumvent both problems (Nicenboim and Vasishth, 2016). Moreover, the *a posteriori* probability of effects given by Bayesian models is particularly adequate to investigate a hierarchy of effects as intended in the present study.

TABLE II. Cutoff-values (min./max, in Hz) for formants (F1, F2) used for excluding vowels from the analysis.

	F1		F2	
Vowel	Minimum	Maximum	Minimum	Maximum
/i/	200	700	1500	3200
/a/	400	1100	900	2200
/u/	200	750	450	1700

Bayesian linear regression models were fit to each acoustic measure [using STAN, Carpenter *et al.* (2017)], after converting them to z-scores (for detailed information see supplemental materials[1]). The dependent variables vowel duration and variability, f0 and variability as well as vowel clarity and variability were predicted by interactive stimulation (arousing vs calm), vocalization type (spoken vs sung), and infant-directness (ID vs non-ID). Mother was used as a grouping variable. The models comprised only main effects as the aim was (1) to identify the hierarchy of effects on the acoustics that are strongest for each predictor, and (2) to keep the models comparable across the acoustic variables. Following Jones and Brandt (2019), in the absence of a strong prior assumption, a generic weakly informative prior was set across β parameters (a normal distribution centered on zero with a SD of 1), this prior being overwhelmed by the large number of observations (849 to 2894 observations for each combination of predictors). A sensitivity analysis (see supplemental materiel[1]) revealed that adjusting the prior had indeed almost no effect on the posterior probabilities. Models fitted successfully for each dependent variable, as shown by R-hat statistics not deviating from the ideal value 1 by more than 0.1, a number of effective samples higher than 40 [5 times the number of chains after splitting, following Gelman *et al.* (2013)], and visual inspection of the trace plots.

III. RESULTS

Figure 1 shows the probability distribution (mean β values and lower and upper bounds of credible intervals) for each predictor and dependent variable. An estimate (β value) of zero indicates a null probability for the effect, while a larger estimate indicates a larger probability. Although such results cannot be interpreted in the same way as those obtained in frequentist analyses [see Vasishth *et al.* (2018), for a discussion], a parallel can be drawn between the interpretation of results in Bayesian analyses: a probability distribution with 95% credible interval not overlapping the zero value can be regarded as reflecting a relevant effect, while compared mean values of estimates may be interpreted as reflecting effect sizes with the standard deviation of the variable of interest as unit.

The estimated effects (see Fig. 1 for mean β values and lower and upper bounds of credible intervals) show that interactive stimulation (arousing vs calm interactions) and vocalization type (sung vs spoken) of codified mother-infant interactions had clear acoustic signatures. As expected, the estimated effect of interactive stimulation, that is, arousing vs calm interactions, is strongest in the pitch dimension (height and variability of f0). As can be seen in Fig. 2, both f0 and variability are considerably higher in arousing than calm interactions. We also found a smaller effect of durational variability of vowels, being more important in calm than in arousing interactions. Importantly, the estimated effects on vowel durations or vowel clarity/variability are negligible in this dimension. Vocalization type, that is differences between speaking and singing, has its most

https://doi.org/10.1121/10.0008977

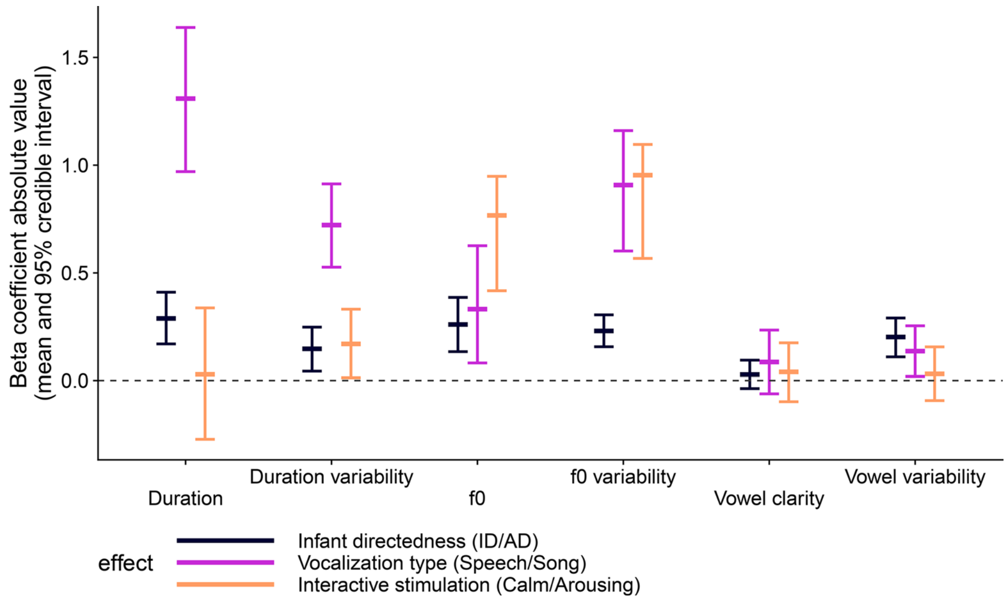

FIG. 1. (Color online) Probability distribution (estimates and 95% credible interval, *y* axis) for each predictor ("effect") and dependent acoustic variable (*x* axis), converted in absolute values to enable comparisons among predictors and variables. Note that an estimate (mean β coefficient) with zero value corresponds to a null probability for the effect, a credible interval overlapping with zero therefore indicates a null effect. Larger effects are associated with larger estimates. The directionality for each variable and predictor can be seen in Fig. 2.

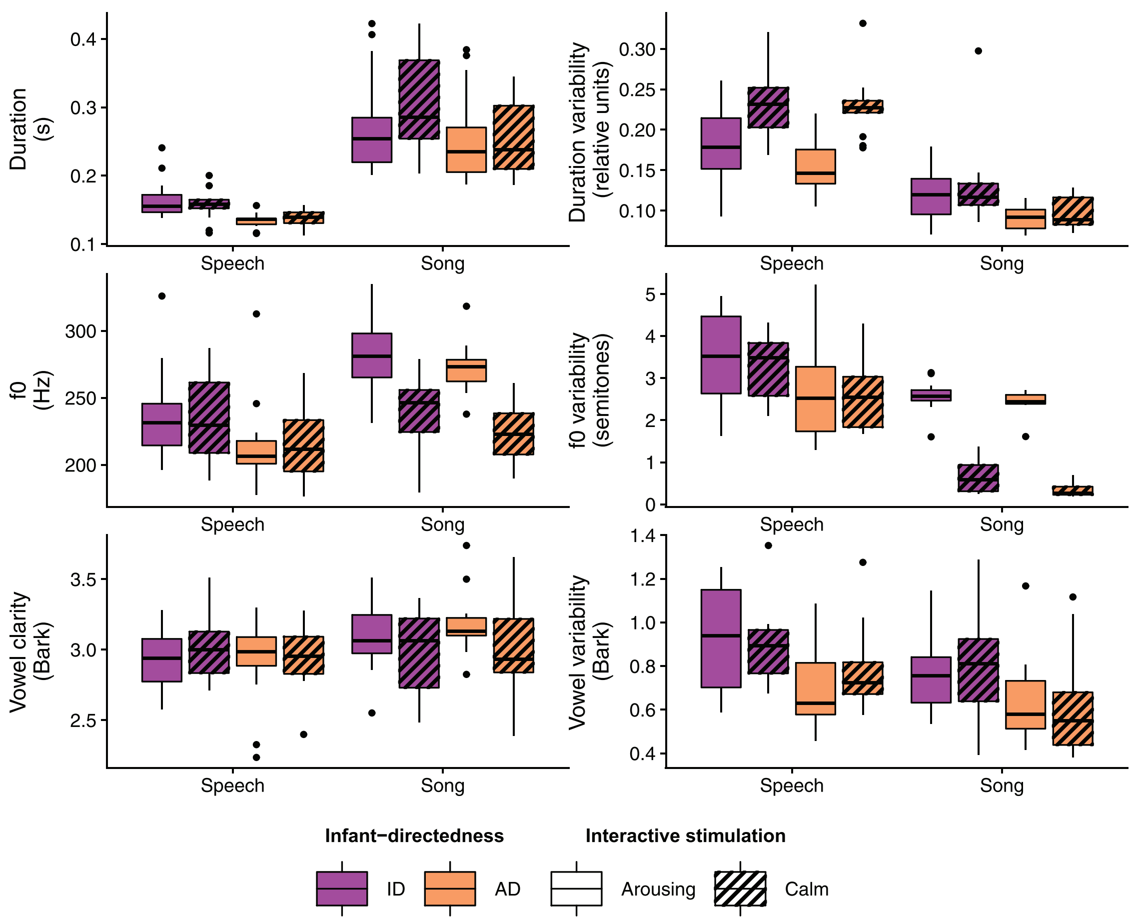

FIG. 2. (Color online) Boxplots for each acoustic measure grouped by condition: infant-directedness (ID pink/AD orange), interactive stimulation (arousing: plain boxes/calm: striped boxes), and vocalization type (spoken/sung on the *x* axis of each panel). The left panels display the values for duration, f0, and vowel clarity, while the right panels display associated variability measures.

important impact on vowel duration (speech < song, Fig. 2). Yet, duration variability (speech > song), f0 variability (speech > song), and f0 (speech > song) as well as vowel variability (speech > song) are also affected, although to a smaller extent. Again, estimated effects on vowel clarity were close to zero. Infant-directedness (ID vs AD), in contrast, was less characteristic in its acoustic signature. All acoustic variables except vowel clarity were likely to vary as a function of infant-directedness in this sample (with all variables having the same direction ID > non-ID, see Fig. 2). However, the magnitude of these estimated effects is modest to small relative to the magnitudes of estimates found for the other communicative dimensions.

IV. DISCUSSION

The present findings reveal prominent acoustic traits ("acoustic signatures") for the communicative dimensions interactive stimulation and vocalization type in codified mother-infant interactions. The pitch dimension—pitch height (f0) and variability—very clearly shapes the acoustic signature of interactive stimulation, measured as a contrast between calm and arousing codified interactions. This finding is in line with previous results on ID spontaneous speech showing an important contribution of variations in the pitch dimension to convey parental messages [e.g., warning, appraisal, soothing etc., Fernald (1989) and Papoušek *et al.* (1991)]. Higher f0, more arousing in both music and speech, is also associated with happiness and positive emotions of the speaker [e.g., Juslin and Laukka (2003)]. Infants are particularly attracted to and aroused by "happy sounds" conveyed by pitch and spectral information (Cirelli *et al.*, 2020; Corbeil *et al.*, 2013). They are also sensitive to the matching between pitch information and the associated interactive context. For example, infants aged 6 to 7 months prefer lower-pitched versions of lullabies compared to higher-pitched versions of playsongs (Tsang and Conrad, 2010; Volkova *et al.*, 2006), although they generally prefer higher pitch in speech and song (Trainor and Zacharias, 1998). Hence, at 6 months, infants could learn from codified interactions that pitch form is highly correlated with paralinguistic functions, thereby laying the grounds for later learning of affective and social dimensions of communication.

Durational variability of vowels was an additional, but less important trait in the acoustic signature of interactive stimulation, with a smaller effect compared to the f0 dimension. However, this effect and its somewhat surprising directionality (calm > arousing) may have been biased by the less rhythmically regular and thereby, more temporally variable structure of calm story reading compared to the metrical or musical beat structure of the other performances.

Vowel duration played an eminent role in the acoustic signature of vocalization type, speech, or song, during the interaction. Long vowel durations are characteristic of singing, as melodic and rhythmic structure is encoded in the most sonorant portions of the signal (Sundberg, 1987, 1989). Consequently, durational vowel-consonant ratios are very indicative of speech vs song [5:1 in song, 1:1 in speech, estimations by Eckardt (1999)]. Variability of vowel durations as well as f0 and variability appeared as additional traits characterizing the acoustic signatures of speech vs song. However, these latter effects and their directions (in the present German corpus: speech > song) should be regarded with caution as they are prone to heavily vary with cultural and individual use of the codified interaction. For example, cross-linguistically, the variability of vowel durations in speech utterances depends on the language-specific prosody, the interlocutor and speech context (Cohen Priva *et al.*, 2017; Klatt, 1976; Nolan and Asu, 2009). The variability of sung vowel durations and f0 is influenced by the metrical and melodic musical structure which is culture-dependent [e.g., London (2004) and Soley and Hannon (2010)]. Although across continents, anthropologists and literary researchers have suggested that the codified repertoire of children is (at least metrically) less variable across cultures than the adult repertoire (Chukovsky, 1963; Burling, 1966), large-scale cross-linguistic and cross–cultural research [e.g., Mehr *et al.* (2018)] might shed light on the actual amount of this variability.

As to spectral properties, it is noteworthy that vowel variability, but not clarity of focal vowels (/a, i, u/) was likely to be predicted by vocalization type (speech > song) in codified mother-infant interactions, although the effect was small. When comparing sung (classical Western operatic singing) and spoken vowels in non-ID interactions, smaller vowel space and reduced intra-category variability have been pointed out as typical of singing (Bradley, 2018), along other special features, such as "the singer's formant" and reduced vowel intelligibility [e.g., Sundberg (1982)]. Singing styles approaching spoken language, however, may attenuate some of these differences [e.g., vowel intelligibility, Sundberg and Romedahl (2009)]. The present findings suggest that, similarly, ID singing may provide more stable realizations of spectral vowel properties than ID speech, a feature that could be further investigated in its function for lexical and phonological processing in infants.

Overall, the results on the type of vocalization, speech, or song, indicate that codified interactions could help infants to learn that acoustic temporal structure is prominently associated with the function of marking differences between spoken and musical registers. This may also foster their capacity to interpret finer temporal differences in linguistic and musical subregisters, such as linguistic styles [clear speech vs colloquial speech, Ferguson and Kewley-Port (2007)] or musical styles [e.g., Savage *et al.* (2015)], later on.

Finally, infant-directedness affected more acoustic parameters in codified interactions than the other dimensions under investigation and no dominant acoustic parameter could be pointed out. However, similar traits were present as reported for spontaneous ID speech, like longer vowel durations/slower tempo [e.g., Fernald *et al.* (1989) and McMurray *et al.* (2013)], higher temporal variability [e.g., Falk and Kello (2017)], higher and more variable f0 (Fernald *et al.*, 1989; Fernald and Kuhl, 1987), and more within-category vowel variability (Martin *et al.*, 2015; McMurray *et al.*, 2013) compared to AD speech. This result suggests that the acoustic signature of infant-directedness is potentially multidimensional, that is, the infant may recover this information from multiple cues. The relatively small estimates for the effects may derive from the fact that codified interactions are *per se* infant-directed, that is, stemming from a historically and culturally evolved infant-directed repertoire, even when performed for an adult. Therefore, the AD performances in the present corpus may not have been as markedly different as are ID vs AD speech during spontaneous conversations [e.g., Sulpizio *et al.* (2018) and Weirich and Simpson (2019)].

Interestingly, unlike previous studies comparing ID and AD speech [e.g., Kalashnikova and Burnham (2018) and Weirich and Simpson (2019)], there were no vowel clarity effects (i.e., more distant vowels /a/, /i/, /u/) in the present corpus. There are a number of possible reasons for this. First, the ritualized and highly repeated character of dyadic codified interactions [see Vlismas *et al.* (2013) for similar results] may particularly enhance reciprocal engagement of mother and infant in the interaction (Margulis 2014), but not linguistic meaning or form. Second, low infant age (6 month) in the study could have counteracted clarity effects as vowel clarity in mothers' speech seems to peak during infants' critical phase for word learning which starts around a year of life [e.g., Bernstein Ratner (1984); see also Han *et al.* (2018), for similar results on lexical tone]. Third, even in spontaneous ID interactions, vowel clarity effects are not always reported across studies, although they are more replicated with focal vowels (/a, i, u/) than with other vowels (Burnham *et al*, 2002; Kuhl *et al.*, 1997, McMurray *et al.*, 2013). For example, Cristia and Seidl (2014) and Martin *et al.* (2015) found higher within-category variability of vowels, but not larger vowel space. The exact function of vowel spectral properties in ID speech are currently debated. McMurray *et al.* (2013) suggest that greater variability of vowels in ID speech could help infants to establish a phonetic/phonological system. Some researchers underline that higher ID vowel clarity derives from articulatory properties of ID speech and hence, link the function to speech production (Best *et al.*, 2016; Kalashnikova *et al.*, 2017), while others highlight the perceptual role of acoustically clearer vowels for infant language learners (Fernald, 2000; Kuhl *et al.*, 1997).

V. CONCLUSION

The present study shows that specific acoustic traits of codified mother-infant interactions highlight specific form-function associations, in particular linked to socio-affective information and differences between musical and linguistic repertoire. On these grounds, future studies could pinpoint differential benefits of codified vs spontaneous interactions in early childcare, and examine, cross-culturally, how infants actually extract form-function associations from codified interactions.

ACKNOWLEDGMENTS

This work was supported by an MCRI-SSHRC Canada research grant (www.airsplace.ca) as well as by the LABEX EFL program (Grant No. ANR-10-LABX 0083). We thank Anne Zorn for help with data collection.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0008977 for SuppPub1.pdf for stimuli material, SuppPub2.txt for the full data set, SuppPub3.pdf for the statistics' code, and SuppPub4.pdf for a figure showing the outcomes of the sensitivity analysis.

Audibert, N., and Falk, S. (**2018**). "Vowel space and f0 characteristics of infant-directed singing and speech," in *Proceedings of the 9th International Conference on Speech Prosody 2018*, pp. 153–157.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**(4), 390–412.

Bergeson, T., and Trehub, S. E. (**2002**). "Absolute pitch and tempo in mothers' songs to infants," Psychol. Sci. **13**, 72–75.

Bernstein Ratner, N. (**1984**). "Patterns of vowel modification in mother-child speech," J. Child Lang. **11**, 557–578.

Best, C. T., Goldstein, L. M., Nam, H., and Tyler, M. D. (**2016**). "Articulating what infants attune to in native speech," Ecol. Psychol. **28**(4), 216–261.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phon. Sci. **17**(1193), 97–110.

Bradley, E. D. (**2018**). "A comparison of the acoustic vowel spaces of speech and song," Ling Res **35**(2), 381–394.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**(3), 255–272.

Burling, R. (**1966**). "The metrics of children's verse: A cross-linguistic study," Am. Anthropol. **68**, 1418–1441.

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

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (**2017**). "Stan: A probabilistic programming language," J. Stat. Softw. **76**(1), 1–32.

Chukovsky, K. (**1963**). *From Two to Five [Ot Dvux Do Pjatij]* (University of California Press, Oakland, CA).

Cirelli, L. K., Jurewicz, Z. B., and Trehub, S. E. (**2020**). "Effects of maternal singing style on mother–infant arousal and behavior," J. Cogn. Neurosci. **32**(7), 1213–1220.

Cohen Priva, U., Edelist, L., and Gleason, E. (**2017**). "Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline," J. Acoust. Soc. Am. **141**(5), 2989–2996.

Conrad, N. J., Walsh, J., Allen, J. M., and Tsang, C. D. (**2011**). "Examining infants' preferences for tempo in lullabies and playsongs," Can. J. Exp. Psychol. **65**, 168–172.

Cooper, R. P., and Aslin, R. N. (**1990**). "Preference for infant-directed speech in the first month after birth," Child Dev. **61**(5), 1584–1595.

Corbeil, M., Trehub, S. E., and Peretz, I. (**2013**). "Speech vs. singing: Infants choose happier sounds," Front. Psychol. **4**, 1–11.

Corbeil, M., Trehub, S. E., and Peretz, I. (**2016**). "Singing delays the onset of infant distress," Infancy **21**(3), 373–391.

Costa-Giomi, E., and Ilari, B. (**2014**). "Infants' preferential attention to sung and spoken stimuli," J. Res. Music Educ. **62**, 188–194.

Cristia, A., and Seidl, A. (**2014**). "The hyperarticulation hypothesis of infant-directed speech," J. Child Lang. **41**(4), 913–934.

De Looze, C., and 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, pp. 135–138.

Eckardt, F. (**1999**). *Singen und Sprechen im Vergleich artikulatorischer Bewegungen* (Singing and Speaking—A Comparison of Articulatory Movements) (THIASOS Musikvlg, Berlin).

Falk, S. (**2009**). *Musik und Sprachprosodie—kindgerichtetes Singen im frühen Spracherwerb* (Music and speech prosody—Infant-directed singing in early language acquisition) (de Gruyter, Berlin).

Falk, S. (**2011**). "Melodic vs. intonational coding of communicative functions—A comparison of tonal contours in infant-directed song and speech," Psychomusicology **21**, 53–68.

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

Farran, L. K., Lee, C.-C., Yoo, H., and Oller, D. K. (**2016**). "Cross-cultural register differences in infant-directed speech: An initial study," PLoS One **11**(3), e0151518.

Fava, E., Hull, R., Baumbauer, K., and Bortfeld, H. (**2014**). "Hemodynamic responses to speech and music in preverbal infants," Child Neuropsychology **20**, 430–438.

Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. **50**, 1241–1255.

Fernald, A. (**1989**). "Intonation and communicative content in mothers' speech to infants: Is the melody the message?," Child Dev. **60**, 1497–1510.

Fernald, A. (**2000**). "Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition," Phonetica **57**(2–4), 242–245.

Fernald, A., and Kuhl, P. (**1987**). "Acoustic determinants of infant preference for motherese speech," Infant Behav. Dev. **10**(3), 279–293.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., Boysson-Bardies, B. de., and Fukui, I. (**1989**). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," J. Child Lang. **16**(3), 477–501.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (**2013**). *Bayesian Data Analysis*, 3rd ed. (Chapman and Hall/CRC, Boca Raton, FL).

Han, M., de Jong, N. H., and Kager, R. (**2018**). "Lexical tones in Mandarin Chinese infant-directed speech: Age-related changes in the second year of life," Front Psychol **9**, 434.

Jones, S. D., and Brandt, S. (**2019**). "Neighborhood density and word production in delayed and advanced learners," J. Speech Lang. Hear. Res. **62**(8), 2847–2854.

Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Psychol. Bull. **129**(5), 770–814.

Kalashnikova, M., and Burnham, D. (**2018**). "Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions," J. Child. Lang. **45**(5), 1035–1053.

Kalashnikova, M., Carignan, C., and Burnham, D. (**2017**). "The origins of babytalk: Smiling, teaching or social convergence?," R. Soc. Open Sci. **4**, 170306.

Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**(5), 1208–1221.

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M., and Fellman, V. (**2010**). "Hemodynamic responses to speech and music in newborn infants," Hum. Brain Mapp. **31**, 595–603.

Kuhl, P. K., Andruški, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**(5326), 684–686.

London, J. (**2004**). *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford University Press, Oxford).

Malloch, S. N. (**1999**). "Mothers and infants and communicative musicality," Music Sci. **3**(1), 29–57.

Margulis, E. H. (**2014**). "Verbatim repetition and musical engagement," Psychomusicology **24**(2), 157–163.

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

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (**2015**). "Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis," Psychol. Sci. **26**(3), 341–347.

Masataka, N. (**1999**). "Preference for infant-directed singing in 2-day-old hearing infants of deaf parents," Dev. Psychol. **35**(4), 1001–1005.

Maurer, D. (**2016**). *Acoustics of the Vowel-Preliminaries* (Peter Lang International Academic Publishers, Pieterlen, Switzerland), pp. 47–51.

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

Mehr, S. A., Singh, M., York, H., Glowacki, L., and Krasnow, M. M. (**2018**). "Form and function in human song," Curr. Biol. **28**(3), 356–368.

Mehr, S. A., Song, L. A., and Spelke, E. S. (**2016**). "For 5- month-old infants, melodies are social," Psychol. Sci. **27**, 486–501.

Montag, J. L., Jones, M. N., and Smith, L. B. (**2015**). "The words children hear: Picture books and the statistics for language learning," Psychol. Sci. **26**(9), 1489–1496.

Nakata, T., and Trehub, S. E. (**2004**). "Infants' responsiveness to maternal speech and singing," Infant Behav. Dev. **27**(4), 455–464.

Narayan, C. R., and McDermott, L. C. (**2016**). "Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations," J. Acoust. Soc. Am. **139**(3), 1272–1281.

Nicenboim, B., and Vasishth, S. (**2016**). "Statistical methods for linguistic research: Foundational Ideas—Part II," Lang. Ling. Compass **10**(11), 591–613.

Nolan, F., and Asu, E. L. (**2009**). "The pairwise variability index and coexisting rhythms in language," Phonetica **66**(1-2), 64–77.

Opie, I., and Opie, P. (**1997**). *The Oxford Dictionary of Nursery Rhymes*, 2nd ed. (Clarendon, Oxford).

Papoušek, M., Papoušek, H., and Symmes, D. (**1991**). "The meanings of melodies in motherese in tone and stress languages," Infant Behav. Dev. **14**, 415–440.

Pätzold, M., and Simpson, A. P. (**1997**). "Acoustic analysis of German vowels in the Kiel Corpus of Read Speech," Arbeitsberichte des Instituts Phonetik und digitale Sprachverarbeitung Universität Kiel **32**, 215–247.

Phillips-Silver, J., and Keller, P. E. (**2012**). "Searching for roots of entrainment and joint action in early musical interactions," Front Hum. Neurosci. **6**, 26.

Piazza, E. A., Iordan, M. C., and Lew-Williams, C. (**2017**). "Mothers consistently alter their unique vocal fingerprints when communicating with infants," Curr. Biol. **27**(20), 3162–3167.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (**2014**). "Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development," Dev. Sci. **17**(6), 880–891.

Rock, A. M. L., Trainor, L. J., and Addison, T. L. (**1999**). "Distinctive messages in infant-directed lullabies and play songs," Dev. Psychol. **35**, 527–534.

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (**2015**). "Statistical universals reveal the structures and functions of human music," Proc. Natl. Acad. Sci. U.S.A. **112**(29), 8987–8992.

Schiel, F. (**1999**). "Automatic phonetic transcription of non-prompted speech," Proc. Int. Congr. Phon. Sci. **1999**, 607–610.

Soley, G., and Hannon, E. E. (**2010**). "Infants prefer the musical meter of their own culture: A cross-cultural comparison," Dev Psychol **46**(1), 286–292.

Sundberg, J. (**1982**). "Perception of singing," in *The Psychology of Music*, edited by D. Deutsch (Academic Press, New York), pp. 59–98.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb, IL).

Sundberg, J. (**1989**). "Synthesis of singing by rule," in *Current Directions in Computer Music Research*, edited by M. V. Mathews and J. R. Pierce (MIT Press, Cambridge, MA), pp. 45–55.

Sundberg, J., and Romedahl, C. (**2009**). "Text intelligibility and the singer's formant—A relationship?," J. Voice **23**, 539–545.

Sulpizio, S., Kuroda, K., Dalsasso, M., Asakawa, T., Bornstein, M. H., Doi, H., Esposito, G., and Shinohara, K. (**2018**). "Discriminating between mothers' infant- and adult-directed speech: Cross-linguistic generalizability from Japanese to Italian and German," Neurosci. Res. **133**, 21–27.

Trainor, L. J., Clark, E. D., Huntley, A., and Adams, B. A. (**1997**). "The acoustic basis of preferences for infant-directed singing," Infant Behav. Dev. **20**, 383–396.

Trainor, L. J., and Zacharias, C. A. (**1998**). "Infants prefer higher-pitched singing," Infant Behav. Dev. **21**, 799–805.x.

Trehub, S. E., and Gudmundsdottir, H. R. (**2019**). "Mothers as singing mentors for infants," in *Oxford Handbook of Singing*, edited by G. F. Welch, D. M. Howard, and J. Nix (Oxford University Press, Oxford), pp.455–469.

Trehub, S. E., and Trainor, L. (**1998**). “Singing to infants: Lullabies and play songs,” Adv. Infancy Res. **12**, 43–78.

Tsang, C. D., and Conrad, N. J. (**2010**). “Does the message matter? The effect of song type on infants’ pitch preferences for lullabies and play-songs,” Infant Behav. Dev. **33**, 96–100.

Tsang, C. D., Falk, S., and Hessel, A. (**2017**). “Infants prefer infant-directed song over speech,” Child Dev. **88**(4), 1207–1215.

Van Puyvelde, M., Vanfleteren, P., Loots, G., Deschuyffeleer, S., Vinck, B., Jacquet, W., and Verheist, W. (**2010**). “Tonal synchrony in mother-infant interaction based on harmonic and pentatonic series,” Infant Behav. Dev. **33**, 387–400.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (**2018**). “Bayesian data analysis in the phonetic sciences: A tutorial introduction,” J. Phon. **71**, 147–161.

Vlismas, W., Malloch, S., and Burnham, D. (**2013**). “The effects of music and movement on mother–infant interactions,” Early Child Dev Care **183**(11), 1669–1688.

Volkova, A., Trehub, S. E., and Schellenberg, E. G. (**2006**). “Infants’ memory for musical performances,” Dev Sci **9**(6), 583–589.

Weirich, M., and Simpson, A. (**2019**). “Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech,” J. Speech Lang. Hear. Res. **62**(11), 4001–4014.