Title
Exploring the role of music in language development: Caregiver song negatively correlates with infant speech-related vocalization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Mai, Jeffrey
VPS, Ritwika
Schneider, Sara
et al.

Publication Date
2026

Exploring the role of music in language development: Caregiver song negatively correlates with infant speech-related vocalizations

Jeffrey Mai (maijiahao@g.ucla.edu)[1], V.P.S. Ritwika[1], Sara Schneider[2], Christopher T. Kello[3], Gregory A. Bryant[1], Anne S. Warlaumont[1,4]
[1]Department of Communication, University of California, Los Angeles
[2]Nicotine and Cannabis Policy Center, Health Sciences Research Institute, University of California, Merced
[3]Cognitive and Information Sciences, University of California, Merced
[4]Department of Psychology, University of California, Los Angeles

Abstract

Caregiver infant-directed (ID) speech and infant vocal productions are both important for speech-language development. Although music experience is associated with language abilities at later ages, the role of music in infant language development is unclear. We used manual annotation to obtain quantities of different infant vocalization types and quantities of caregiver speech and song during 5-minute excerpts from daylong audio recordings of 3-, 6-, 9- and 18-month-olds. Excerpts in which caregivers produced more infant-directed musical vocalizations (such as singing or humming) also contained fewer speech-related infant vocalizations. This could be driven by musical input causing infants to be quieter as they attend auditorily and/or by adults singing or humming during times when infants are less vocal. We also found a negative association between non-infant-directed adult speech and infant speech-related vocalization, positive associations between quantities of infant-directed speech and infant crying and laughter, and age trends consistent with prior reports.

Keywords: language development; infancy; vocalization; singing; music; infant-directed speech

Introduction

Music training has potentially large benefits for language development, including tuning the brain for basic auditory skills (Kraus & Chandrasekaran, 2010) and speech discrimination in noise (Strait et al., 2013), and studies have found that phonological awareness is associated with musical skills such as tempo and rhythm perception (Anvari et al., 2002; Hutchins, 2018; Politimou et al., 2019). The mere presence of music in the home environment has been shown to predict word comprehension ability (Papadimitriou et al., 2021), and musical interventions can enhance neural processing of temporal structure in speech in both typically developing infants (Zhao & Kuhl, 2016) and those at risk for dyslexia (Paula et al., 2023).

Much of the research on relationships between exposure to music and subsequent language learning has been conducted with preschool-aged children and older. While interactive musical training starting as early as 6 months has been found to enhance music ability as well as communicative and other social skills (Gerry et al., 2012). Little research has focused on the role of music in everyday vocal interactions. Exploring potential associations between musical adult vocalizations and infants' pre-speech and early-speech vocalizations could shed light on when and how associations between music and language get their start.

Infant vocalizations lay the foundation for later phonological skills in speech production (Oller, 2000). For infants to communicate their needs and develop motor and cognitive speech skill foundations, they must first explore their vocal capabilities through a variety of vocalization types, pre-speech vocalization, cries, and laughs. They often perform this exploration in the context of infant-caregiver vocal interactions, as well as during independent play (Long et al., 2022; Yoo et al., 2021).

When caregivers speak to infants, including in such interactive contexts, they tend to modify their speech compared to when they speak to other adults. Adult vocalizations that are specifically directed to infants are referred to as infant-directed speech or infant-directed song (Cristia, 2013; Spinelli et al., 2017). These modifications include higher pitch, greater pitch variability, extended word length, and exaggerated prosodic contours (Cristia, 2013; Spinelli et all., 2017). These infant-directed speech (IDS) modifications are thought to make speech more perceptible, thereby making language more learnable (Kitamura et al., 2001; Thiessen et al., 2005). Indeed, the quantity of IDS produced by adults has been shown to predict infant language abilities (Rollins, 2003; Weisleder & Fernald, 2013).

Live singing is often prominent in daylong child-centered audio recordings (Mendoza & Fausey, 2021; Neer et al., 2025). Compared to speech, these musical stimuli may be processed and attended to differently by infants. Infants can distinguish between infant-directed speech and infant-directed song as early as 7 months and engage with the song based on the language (Tsang et al., 2017). One study found that while speech stimuli typically receive relatively longer infant vocal responses, music stimuli often receive short bursts of vocal responses (Reigado et al., 2011).

In the current study, we addressed the knowledge gap regarding how infants' real-world experiences of musical and non-musical adult input—both infant-directed and other-directed (OD)—related to infants' own vocal productions. Utilizing 5-minute high-infant-volubility samples from daylong recordings from a longitudinal study of infants at 3, 6, 9, and 18 months, we assessed how much live vocal singing and speech infants are exposed to, whether that input

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

was infant-directed or other-directed, and how the quantities of those different types of adult vocalizations related to quantities of infant speech-related vocalizations (both canonical and non-canonical), cry, and laugh. If infant-directed song and infant-directed speech are both positively associated with periods of increased speech-related infant vocal production, especially canonical utterances, which are pre-speech vocalizations that resemble adult speech that contain both a vowel-consonant combination and speech-like transitions, it would suggest that both types of preverbal infant-caregiver interactions contribute to language development by promoting speech and pre-speech vocalizations. However, infant-directed singing could also conceivably be negatively associated with infant protophones, which would suggest that infant-directed song and infant-directed speech may play different roles in infants' prelinguistic development.

Method

Participants

The dataset is a corpus of daylong home audio recordings and questionnaire data collected longitudinally at 3, 6, 9, and 18 months from 54 infants living in the San Joaquin Valley (Ritwika et al., 2020, 2025) recruited by word of mouth, hospital tours, and flyers. Participants included families with varying home languages, predominantly English and Spanish. Monetary compensation was provided based on completion of each recording. SES of caregivers varied: seven households reported annual income less than \$30,000, seven reported between \$30,000–60,000, six reported between \$60,000–90,000, and nine reported over \$90,000. Of the 54 infants, 17 were reported non-Hispanic White, 13 Hispanic/Latino, 13 Hispanic/Latino - White, 2 Hispanic/Latino - Native American, 2 Black/African American, 4 White & Asian, and 3 Unreported. The highest level of education within the families also varied: 2 completed some high school, 3 completed high school or received a GED, 10 completed some college, 3 completed an associate's degree, 18 completed a bachelor's degree, 11 completed a master's degree, and 7 completed a doctoral degree. This study utilized a subset of the corpus, consisting of 420 5-minute sections selected from 148 daylong recordings, including 43 recordings at 3 months, 35 at 6 months, 37 at 9 months, and 33 at 18 months. This included 45 infants, with 30 infants represented in all four age groups.

Data Collection

Daylong recordings were obtained using Language ENvironment Analysis (LENA; lena.org) recorders that were lent to participating families. Caregivers were provided with written instructions to turn on the recorders from approximately 8AM to 7PM, collecting a minimum of 10 hours of data per recording. If needed, caregivers were allowed to pause the recording for up to one total hour during the day if there were moments containing sensitive information. If they were unable to pause in advance, caregivers were also allowed to request cuts to be made to omit sections after the recordings were uploaded. Caregivers also completed a few paper-based questionnaires on each recording day. The questionnaire data were not analyzed for this study.

Selection of 5-Minute Segments

We used the LENA Pro system's ADEX tool to tabulate the counts of the LENA system's automatically generated sound source labels at the 5-minute-level. These source labels include the target child, other children, and adult male and female speakers, TV/electronic sounds, noise, and overlap (Gilkerson & Richards, 2020). This information was used to identify the three highest infant volubility portions within each daylong recording. If either the second or third highest volubility section were within 30 minutes to either the first or second highest volubility section, respectively, they were to be replaced with the next-highest volubility 5-minute section as needed. This was to ensure variation in the episodes and contexts represented in the manually annotated dataset. However, in some cases this replacement did not get implemented due to human error, so that in 17 of the daylong recordings there are 5-minute sections less than 30 minutes apart (Ritwika et al., 2025). High volubility sections were used instead of random selection to ensure infants were awake and active and to maximize the number of infant vocalizations—with pure random sampling we would have had to perform more human coding to obtain sufficient numbers of infant vocalizations, especially of rarer types.

The 5-minute high-volubility sections were randomly assigned to trained human listeners. For all annotations except the coding of which adult vocalizations were musical, this was done in three phases based on the available data at the time of the start of the pass, since annotation of all behaviors except adult music began prior to the completion of the longitudinal data collection. In a small number of cases, due to human error the portions of recordings that were annotated did not match those assigned; these were excluded from the current study. There were also some cases in which the research assistant did not understand the language(s) being spoken, preventing them from completing the coding. In these cases, either a bilingual researcher who understood the language was called in to perform the coding in place of the original assigned coder or, if there was no such researcher available, the 5-minute sample was excluded. There was also one recording whose three segments were excluded due to a technical issue during the ELAN export process. Finally, six 5-minute segments were dropped during data processing, likely due to human error. In the end, 410 5-minute segments were included in the current study, including 116 at 3 months, 101 at 6 months, 101 at 9 months, and 92 at 18 months.

Data Annotation

Manual annotation was performed in ELAN (ELAN, 2025) by a team of graduate and undergraduate student researchers, including two of the authors (J.M. and S.S.), who received training to follow a coding manual and were given feedback based on annotations of one or more practice files. The onsets and offsets of infant and adult vocalizations were marked based on the following instructions: (1) mark any sound of nontrivial loudness that you think was made by the vocal tract (i.e., larynx, throat, mouth, lips, teeth, nasal passage, etc.), (2) be as accurate as possible when setting utterance boundaries—try to be within tenths of a second, and (3) annotations should extend from the onset to the offset of each sound. Vegetative infant vocalizations (e.g., coughs) were excluded.

Each infant vocalization was coded as canonical speech-related (speech or speech precursor), non-canonical speech-related, cry, or laugh (Oller, 2000). Infant syllables needed to satisfy two criteria to be deemed canonical: (1) a full vowel as the nucleus of the syllable, and (2) speech-like transition between consonant and vowel sounds (Buder et al., 2013). Multisyllabic infant vocalizations (e.g., "mama") were deemed canonical when most syllables were perceived as canonical.

Adult vocalizations were coded as directed to the infant wearing the recorder, to another individual (other-directed), or to an unknown addressee; this decision was based on who the listener perceived to be the intended addressee and listeners utilized the utterance itself as well as any other context they could glean from the audio recording. Unknown addressee utterances were excluded from the current study's analyses. Adult vocalizations were also coded as non-musical, live lyrical musical (i.e., singing), or live non-lyrical musical (e.g., humming, beatboxing). The latter two categories collapsed in the current study into one musical adult vocalization category, which we will refer to as singing since that constituted most of the category, to have sufficient power for statistical analysis. This coding scheme resulted in four adult vocalization categories: infant-directed speech, other-directed (i.e., not-infant-directed) speech, infant-directed song, and other-directed song. All adult speakers/singers in each recording were treated as a single entity because individual speakers were not identified. There were occasional instances of adult vocalization annotations capturing non-adult vocalizations, such as an older sibling.

Reliability

Recordings were randomly selected for reliability coding. The last author (A.W.) was provided the onset and offset markings for 10 randomly selected recordings (each with all three annotated 5-minute sections) and filled in the annotations for all vocalization types. Cohen's kappa for the infant vocal type annotations was 0.516. The only moderate level of agreement may be related to the fuzziness of the distinction between cry and non-cry fussing, among other potential sources of disagreement. There were some instances where coders labelled infant marginal babbles and gooing as canonical, which may account for the high quantity of canonical vocalizations for a few of the three-month-old 5-minute sections. Adult utterance direction had a Cohen's kappa of 0.759 and adult utterance musicality had a Cohen's kappa of 0.854.

Data Analysis

Quantity of a vocalization type was defined as ln(total duration within the 5-minute sample + 1). To aid interpretability, all log durations were *z*-scored before entering them into the analyses and age was scored in months units. Linear mixed effects models were used to predict the total quantity of infant speech-related vocalizations, with infant participant as a random effect, and infant age and the quantity of each adult vocalization type as fixed effects. Following those analyses, the same model was applied within each age group (excluding age as a fixed effect). Analyses predicting the quantity of canonical babbling were performed separately and are only reported for the 9- and 18-month-olds, as only at these ages are most infants expected to have achieved consistent rates of canonical syllable production (Oller, 2000). Finally, using logistic regression, we analyzed whether a 5-minute segment contained any crying or any laughter as a function of each of the four adult vocalization types and infant age, with infant participant as a random effect. All analyses were performed in R using the lme4 and lmerTest packages.

Data tables, code, and full model outputs are available on OSF: https://osf.io/xjkgs

Results

Figure 1 shows the distributions of quantities of the infant vocalization types at the 5-minute-section-level across the four infant ages. There was a mean of 9.90 seconds of canonical infant vocalizations (min = 0.00, max, 82.70, IQR = 14.34), a mean of 39.41 seconds of non-canonical infant vocalizations (min = 0.44, max = 159.40, IQR = 37.83), a mean of 9.010 seconds of infant reflexive cry (min = 0.00, max = 212.65, IQR = 8.76), and 1.17 seconds of infant reflexive laughter (min = 0.00, max = 43.21, IQR = 0.00).

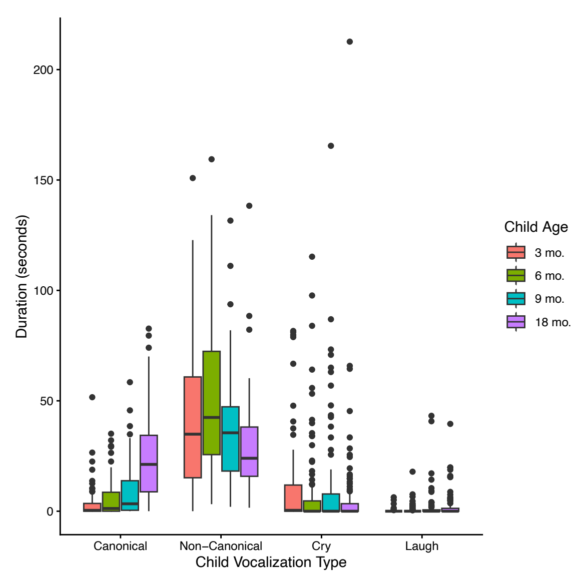

Figure 1: Quantity of each infant vocalization type (total duration in a 5-minute sample) by infant age (months).

Figure 2 shows the distributions of the quantities of adult vocalization types at the 5-minute section level across the four infant ages. There was a mean of 26.14 seconds of infant-directed speech (min = 0.00, max = 152.00, IQR = 35.71), a mean of 23.20 seconds of other-directed speech (min = 0.00, max = 210.71, IQR = 30.22), a mean of 2.86 seconds of infant-directed singing (min = 0.00, max = 126.66, IQR = 0.00), and a mean of 0.50 seconds of other-directed singing (min = 0.00, max = 124.74, IQR = 0.00).

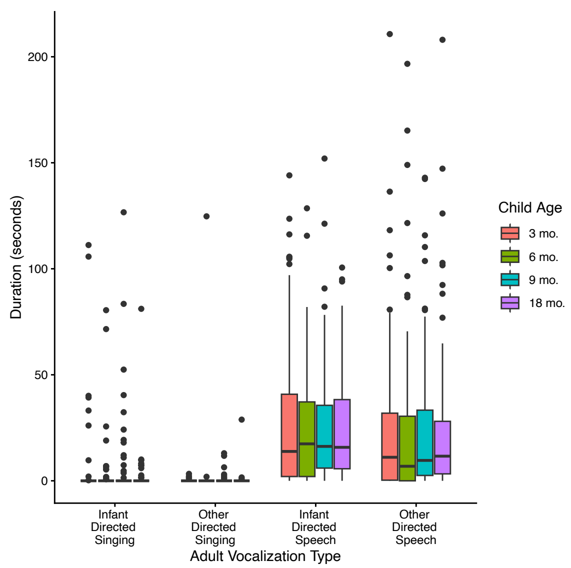

Figure 2: Quantity of each adult vocalization type (total duration in a 5-minute sample) by infant age (months).

Table 1. Mixed effects regression model results predicting infant speech-related vocalization quantity. Boldness indicates statistical significance. Coefficients are standardized (β) except for Age (b).

Predictor	Coefficient	95% C.I.	*p*
Intercept	**-0.27**	**[-0.48, -0.07]**	**.01**
Age (months)	**0.02**	**[0.008, 0.04]**	**.004**
ID Song	**-0.20**	**[-0.30, -0.10]**	**<.001**
OD Song	-0.06	[-0.15, 0.03]	.19
ID Speech	-0.01	[-0.11, 0.09]	.83
OD Speech	**-0.14**	**[-0.23, -0.05]**	**.002**

Marginal R^2 = .085, Conditional R^2 = .282

Table 1 shows the coefficients and marginal and conditional R^2 for the mixed effects model predicting quantity of infant speech-related vocalization (including both canonical and non-canonical utterances) within a 5-minute section based on infant age and the quantities of each of the four adult vocalization types. We found that age was a predictor ($b = 0.02$, $p = .004$), such that the quantity of infant speech-related vocalization increased as infants got older. Adult infant-directed musical vocalization (including singing) was significantly associated with *less* infant speech-related vocalization ($\beta = -0.20$, $p < .001$; see Figure 3) as was adult other-directed speech ($\beta = -0.14$, $p = .002$; see Figure 4). Adult other-directed singing and adult infant-directed speech were not significant predictors.

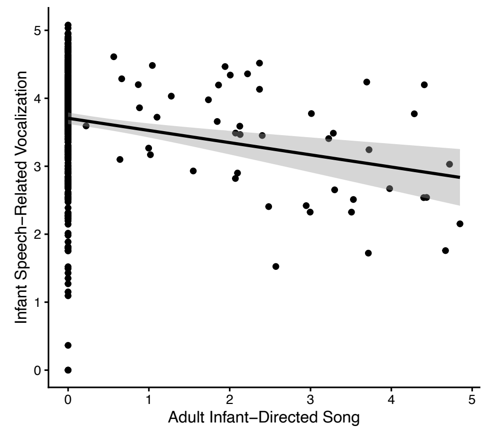

Figure 3: Quantity in log(seconds) of infant speech-related vocalization as a function of quantity of adult infant-directed musical vocalization, for the full dataset containing all age groups. Each point is a 5-minute sample. Here and in Figs. 4 & 5, the line fit and 95% CI are for illustration purposes only; they do not account for random effects, infant age, or quantities of other adult vocalization types.

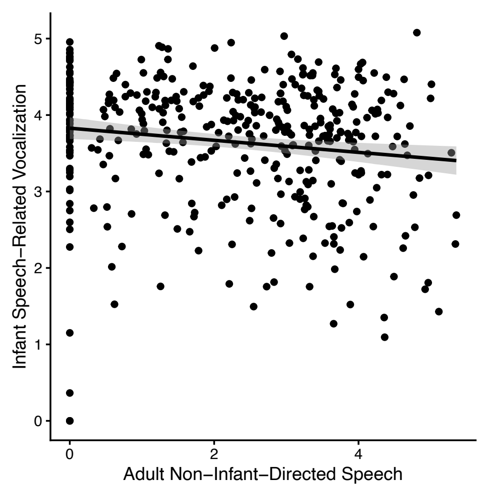

Figure 4: Quantity in log(seconds) of infant speech-related vocalization as a function of quantity of adult other-directed speech, for the full dataset containing all age groups.

Tables 2 and 3 show the results for individual mixed effects regressions predicting quantity of canonical infant vocalization for 9-month-olds and 18-month-olds, respectively. For 9-month-olds, infant-directed singing negatively predicted the quantity of canonical vocalization ($\beta = -0.27$, $p = .02$; see Figure 5). No other adult vocalization types significantly predicted 9-month-old canonical vocalization quantity, and none of the four adult vocalization types significantly predicted 18-month-old canonical vocalization quantity.

Table 2. Mixed effects regression model results predicting 9-month-old infant canonical speech-related vocalization quantity. Boldness indicates statistical significance. Coefficients are standardized (β) except for Age (b).

Predictor	Coefficient	95% C.I.	*p*
Intercept	0.0006	[-0.30, 0.30]	.997
ID Song	**-0.27**	**[-0.49, -0.05]**	**.02**
OD Song	0.002	[-0.22, 0.22]	.98
ID Speech	-0.10	[-0.31, 0.11]	.35
OD Speech	-0.14	[-0.33, 0.05]	.15

Marginal R^2 = .121, Conditional R^2 = .339

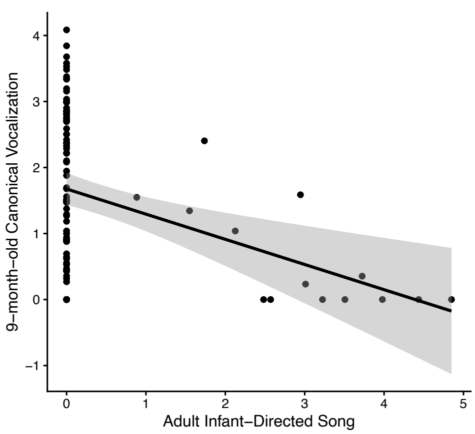

Figure 5: Quantity in log(seconds) of infant canonical vocalization as a function of quantity of adult infant-directed musical vocalization for 9-month-old infants.

Table 3. Mixed effects regression model results predicting 18-month-old infant canonical speech-related vocalization quantity. Boldness indicates statistical significance. Coefficients are standardized (β) except for Age (b).

Predictor	Coefficient	95% C.I.	*p*
Intercept	-0.01	[-0.23, 0.21]	.92
ID Song	0.02	[-0.14, 0.19]	.79
OD Song	0.08	[-0.07, 0.24]	.29
ID Speech	0.04	[-0.15, 0.22]	.69
OD Speech	-0.07	[-0.27, 0.12]	.45

Marginal R^2 = .013, Conditional R^2 = .584

Table 4 shows the coefficients and marginal and conditional R^2 for the logistic mixed effects model predicting whether a 5-minute section contained infant crying. Infant age was associated with less infant crying (b = -0.04, p = .03) and infant-directed speech was positively associated with occurrence of infant crying (β = 0.38, p = .003).

Table 5 shows the coefficients and marginal and conditional R^2 for the logistic mixed effects model predicting whether a 5-minute section contained infant laughter. Infant age was positively associated with presence of infant laughter (b = 0.12, p < .001) as was infant-directed speech (β = 0.53, p < .001).

Table 4. Mixed effects logistic regression model results predicting presence of infant crying. Boldness indicates statistical significance. Coefficients are standardized (β) except for Age (b).

Predictor	Coefficient	95% C.I.	*p*
Intercept	0.14	[-0.32, 0.60]	**.56**
Age (months)	**-0.04**	**[-0.08, -0.004]**	**.03**
ID Song	0.08	[-0.15, 0.31]	.49
OD Song	0.005	[-0.21, 0.23]	.96
ID Speech	**0.38**	**[0.13, 0.63]**	**.003**
OD Speech	-0.15	[-0.38, 0.07]	.18

Marginal R^2 = .053, Conditional R^2 = .206

Table 5. Mixed effects logistic regression model results predicting presence of infant laughter. Coefficients are standardized (β) except for Age (b).

Predictor	Coefficient	95% C.I.	*p*
Intercept	**-2.55**	**[-3.14, -1.96]**	**<.001**
Age (months)	**0.12**	**[0.07, 0.16]**	**<.001**
ID Song	0.06	[-0.19, 0.32]	.63
OD Song	0.005	[-0.25, 0.26]	.97
ID Speech	**0.53**	**[0.22, 0.85]**	**<.001**
OD Speech	0.10	[-0.18, 0.37]	.48

Marginal R^2 = .179, Conditional R^2 = .251

Discussion

We found that when caregivers produced more infant-directed musical vocalizations, infants produced fewer speech-related vocalizations. Moreover, 9-month-olds, who have typically recently established a regular practice of canonical babbling, exhibited a negative association between canonical vocalization quantity and quantity of infant-directed singing. One explanation for this negative association may be that infants become quieter during live vocal singing or humming in comparison to infant-directed speech, attending to the music rather than vocalizing themselves (Reigado et al., 2011). Previous research has found that infants from 6–10 months can perceive the difference between infant-directed speech and song and prefer to listen to the latter (Tsang et al., 2017). Another possibility is that adults may choose to sing during times when infants are producing fewer speech-related vocalizations, as prolonged singing or humming might be inappropriate during times when infants show interest in engaging in dyadic proto-conversation (Duong et al., 2024; Pretzer et al., 2019). A temporally sensitive analysis examining leading and following vocalizations between infants and adults could help shed light on these possibilities.

In any case, the current results suggest that associations between music experience and language skill may not be due to a general tendency rooted early in development for children to be more vocally active in response to live vocal musical input. Instead, associations between music and language learning may relate more to perceptual aspects of language learning, or to more subtle changes in infant vocal practice than the current analysis could detect. This would align with prior findings that early music experiences correlate with greater skill on certain linguistic perception tasks, such as phonological processing, speech discrimination in noise, and other similar tasks (Anvari et al., 2002; Sallat & Jentschke, 2015; Hutchins, 2018). Alternatively, associations between music experience and language learning might be mediated by other factors, such as differences in overall well-being related to music experience. It is also possible that the influences of musical adult vocalization on infant protophone production are indirect or delayed enough in effect that they would be apparent at the full day level but not revealed in the 5-minute-level analyses presented here.

Regarding cry and laughter, IDSong was a positive predictor of both. This aligns with prior findings from manually annotated excerpts from daylong audio recordings of 1-year-olds (Pretzer et al., 2019). Adults were likely engaging in play with infants in response to or triggering laughter (Pisanski et al., 2022) and may have been using speech to try to soothe a crying infant (Pretzer et al., 2019).

It was surprising that IDSpeech was not a significant predictor of infant pre-speech vocalizations. This may be due to the focus on overall quantities within 5-minute high-infant-volubility time periods. Previous research did find significant local temporal contingencies between infant protophones and ID adult vocalizations within high-volubility 5-minute sections of LENA recordings for 1-year-olds (Pretzer et al., 2019).

Limitations and Future Directions

One limitation of the current study was our unit of analysis: log duration during a five-minute high-infant-volubility sample. The decision to focus on these contexts was in part to ensure the infant was awake to attend to the auditory stimuli. It was also motivated by the broad project's goal of capturing larger samples of infant vocal production and possibly also greater numbers of infant-directed behaviors. However, this sampling method has limitations: 1) not capturing the full range of infant auditory experiences; and 2) potentially being biased towards the representation of circumstances surrounding infant vocal activity. As such, future studies using more arbitrary sampling or even full-day analysis would be informative. The latter would be facilitated by tools for automatically coding the different infant and adult vocalization types of interest.

The current analysis was also limited by the approach of running linear mixed effects regressions at the level of the whole 5-minute sample rather than utilizing more temporally precise interaction-based analyses, which may provide more clarity on possible leading-following relationships. For instance, contingent interaction analysis could provide insights on caregiver-infant interactions that may promote infant vocal production or infant behaviors that serve as prompts for adult behaviors (Bornstein & Tamis-LeMonda, 2022; Casillas & Scaff, 2021, Goldstein et al., 2010; Miller et al., 2018; Lopez et al., 2020; Warlaumont et al., 2014). Cross-recurrence quantification analysis (CRQA) is another method that could be utilized to examine patterns of interaction across a range of timescales and investigate leading-following dynamics (Coco et al., 2020; Warlaumont et al., 2014).

Music is a varied category of sound—different types of songs can elicit diverse infant responses. While lullabies are meant to soothe an infant to sleep, vocal play and other types of music and song can occur throughout the day and may be intended to stimulate infants and engage them more actively (Reynolds et al., 2007). Thus, a more detailed coding of both song types and contexts could be helpful to understand the patterns found in the current study as well as potentially uncover patterns that were not apparent when pooling together the various types of songs and contexts represented in the 5-minute samples from daylong audio recordings.

Finally, our study was limited by the audio-only nature of the dataset. The use of video or even infant eye-tracking data in naturalistic settings would potentially allow for more information on how infants attend to musical stimuli and could allow for a better understanding of why live musical adult vocalizations were negatively correlated with infant protophone vocalizations. For instance, video data capturing an infant's gaze during naturalistic home-based interactions would allow for comparison work on infant attention, and adult/infant vocalization type and frequency.

Conclusion

The average day for an infant is filled with rich auditory information. Much of this comes from adult caregivers, including speech and singing, directed to the infant and/or others in the environment. Here, we looked at how different infant vocalization types were associated with infant-directed and other-directed adult speech and song, comparing quantities of vocalization types at the 5-minute level based on real-world infant-centered audio recordings. We were particularly interested in uncovering any potential associations between musical adult vocalizations and the infant speech-related sounds that form the foundation and beginnings of early language development. We found that the quantity of adult musical vocalization was *negatively* associated with the quantity of infant speech-related vocalizations. This finding suggests that associations between music experience and language are unlikely to be rooted in a tendency for musical input to directly stimulate active vocal practice by infants. Future research would benefit from more fine-grained temporal analyses and more detailed coding of music types and contexts.

Acknowledgements

We thank the infants and their families, Gina Pretzer, Sara Mendoza, and the research assistants who contributed to data collection and annotation. Funding was provided by the National Science Foundation (BCS-1529127 and SMA-1539129/1827744), a James S. McDonnell Foundation Fellowship in Understanding Human Cognition (https://doi.org/10.37717/220020507), and UCLA Graduate Research Mentorship and Graduate Summer Research Mentorship fellowships awarded to J.M.

References

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111–130. https://doi.org/10.1016/S0022-0965(02)00124-8

Bornstein, M. H., & Tamis-LeMonda, C. S. (2022). Activities and interactions of mothers and their firstborn infants in the first six months of life: Covariation, stability, continuity, correspondence, and prediction. In *Parenting: Selected writings of Marc H. Bornstein* (pp. 32–49). Routledge.

Buder, E. H., Warlaumont, A. S., Oller, D. K. (2013). An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. Comprehensive In B. Peter & A. MacLeod (Eds.), *Perspectives on child speech development and disorders: Pathways from linguistic theory to clinical practice*, (pp. 103–134). Nova Science Publishers.

Casillas, M., & Scaff, C. (2021). Analyzing contingent interactions in R with chattr. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. Retrieved from https://escholarship.org/uc/item/4rr848x0

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Duong, S., Davis, T. J., Bachman, H. J., Votruba-Drzal, E., & Libertus, M. E. (2024). Exploring dynamic structures of dyadic conversations using categorical cross recurrence quantification analysis: A tutorial. *The Quantitative Methods for Psychology*, *20*(2), 121–136. https://psycnet.apa.org/doi/10.20982/tqmp.20.2.p121

ELAN (Version 7.0) [Computer software]. (2025). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science*, *15*(3), 398–407. https://doi.org/10.1111/j.1467-7687.2012.01142.x

Gilkerson, J., & Richards, J. A. (2020). A Guide to Understanding the Design and Purpose of the LENA® System (Report no. LTR -12). LENA Foundation. https://www.lena.org/wp-content/uploads/2020/07/LTR-12_How_LENA_Works.pdf

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, *15*(4), 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x

Hutchins, S. (2018). Early childhood music training and associated improvements in music and language abilities. *Music Perception: An Interdisciplinary Journal*, *35*(5), 579–593. https://doi.org/10.1525/mp.2018.35.5.579

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425-2438. https://doi.org/10.1121/1.417951

Long, H. L., Ramsay, G., Griebel, U., Bene, E. R., Bowman, D. D., Burkhardt-Reed, M. M., & Oller, D. K. (2022). Perspectives on the origin of language: Infants vocalize most during independent vocal play but produce their most speech-like vocalizations during turn taking. *PLoS ONE*, *17*(12), e0279395. https://doi.org/10.1371/journal.pone.0279395

Lopez, L. D., Walle, E. A., Pretzer, G. M., & Warlaumont, A. S. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *PLoS ONE*, *15*(11), e0242232. https://doi.org/10.1371/journal.pone.0242232

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6), e13122. https://doi.org/10.1111/desc.13122

Miller, J. L., Hurdish, E., & Gros-Louis, J. (2018). Different patterns of sensitivity differentially affect infant attention span. *Infant Behavior and Development*, *53*, 1–4. https://doi.org/10.1016/j.infbeh.2018.10.001

Neer, E. M., Brahmbhatt, A., Walsh, C. R., & Warlaumont, A. S. (2025). Pitch characteristics of real-world infant-directed speech vary with pragmatic context, perceived adult gender, and infant gender. *PLoS ONE*, *20*(6), e0326569. https://doi.org/10.1371/journal.pone.0326569

Paula, V., Vesa, P., Anastasia, G., Anja, T., Laurel J, T., & Teija, K. (2023). Beneficial effects of a music listening intervention on neural speech processing in 0–28-month-old children at risk for dyslexia. *Developmental Science*, *26*(5), e13426. https://doi.org/10.1111/desc.13426

Papadimitriou, A., Smyth, C., Politimou, N., Franco, F., & Stewart, L. (2021). The impact of the home musical environment on infants' language development. *Infant Behavior and Development*, *65*, 101651. https://doi.org/10.1016/j.infbeh.2021.101651

Pisanski, K., Bryant, G. A., Cornec, C., Anikin, A., & Reby, D. (2022). Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution*, *34*(3), 303–321. https://doi.org/10.1080/03949370.2022.2026482

Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in pre-schoolers. *Frontiers in Psychology*, *10*, 948. https://doi.org/10.3389/fpsyg.2019.00948

Pretzer, G. M., Lopez, L. D., Walle, E. A., & Warlaumont, A. S. (2019). Infant-adult vocal interaction dynamics depend on infant vocal type, child-directedness of adult speech, and timeframe. *Infant Behavior and Development*, *57*, 101325. https://doi.org/10.1016/j.infbeh.2019.04.007

Reigado, J., Rocha, A., & Rodrigues, H. (2011). Vocalizations of infants (9–11 months old) in response to musical and linguistic stimuli. *International Journal of Music Education*, *29*(3), 241–255. https://doi.org/10.1177/0255761411408507

Reynolds, A. M., Long, S., & Valerio, W. H. (2007). Language acquisition and music acquisition: Possible parallels. In K. Smithrim & R. Upitis (Eds.), *Listen to their voices, research to practice: A biennial series* (pp. 211–227). Waterloo: Canadian Music Educators Association.

Ritwika, V.P.S., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports*, *10*, 10469. https://doi.org/10.1038/s41598-020-66778-0

Ritwika, V.P.S., Schneider, S., Lopez, L.D., Mai, J., Gopinathan, A., Kello, C.T., & Warlaumont, A.S. (2025). Burstiness and interpersonal foraging between human infants and caregivers in the vocal domain [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2505.01545

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(2), 221–234. https://doi.org/10.1017/S0142716403000110

Sallat, S., & Jentschke, S. (2015). Music perception influences language acquisition: Melodic and rhythmic-melodic perception in children with Specific Language Impairment. *Behavioural Neurology*, *2015*, 606470. https://doi.org/10.1155/2015/606470

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3), 191–201. https://doi.org/10.1016/j.bandl.2012.09.001

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748–767. https://doi.org/10.1111/1467-8624.00313

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants prefer infant-directed song over speech. *Child Development*, *88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Yoo, H., Oller, D. K., & Ha, S. (2021). Early emergence and development of protophones in the first year of life. *Communication Sciences & Disorders*, *26*(1), 1–12. https://doi.org/10.12963/csd.20786

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), 5212–5217. https://doi.org/10.1073/pnas.1603984113