Journal of Child Language (2025), 1–13
doi:10.1017/S0305000925100081

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

Spoken or sung? Examining word learning in child-directed speech and in song

Mackensie Blair, Lindsay Hawtof and Giovanna Morini

Department of Communication Sciences and Disorders, University of Delaware, Newark, DE, USA
Corresponding author: Mackensie Blair; Email: mmblair@udel.edu

(Received 16 September 2024; revised 07 May 2025; accepted 07 May 2025)

Abstract

The present study examines whether presenting words in song versus spoken sentences can lead to differences in word learning in 47–50-month-old children. This work extends previous findings on this topic and evaluates whether the location of pitch changes within the song may contribute to how well the words are learned. Using a Preferential Looking Paradigm, 32 children were taught the names of objects, either in spoken sentences or in a song that followed an unfamiliar melody. In both conditions, the novel word was emphasized by a pitch change. Looking patterns indicated that children learned the names of the novel items better when the words were trained in the spoken sentence compared to the song condition. The findings are discussed in relation to theories of word learning, and how differences in the characteristics between speech and song may relate to variability in how well new words are acquired.

Keywords: word learning; songs; preschoolers

Resumen

Esta investigación examina si la presentación de palabras en canciones en lugar de oraciones habladas puede llevar a diferencias en el aprendizaje de palabras en niños de 47 a 50 meses. Este trabajo amplía hallazgos previos sobre este tema y evalúa si los momentos donde ocurren cambios de tono en la melodía de la canción pueden contribuir a la eficacia con la que se aprenden nuevas palabras. Usando un Paradigma de Mirada Preferencial, a 32 niños se les enseñaron los nombres de varios objetos, tanto en oraciones habladas como en una canción que seguía una melodía desconocida. En ambas condiciones, la palabra nueva se enfatizaba con un cambio de tono. Los patrones de mirada indicaron que los niños aprendieron mejor los nombres de los objetos cuando las palabras fueron entrenadas en la oración hablada en comparación con la versión que incluía las palabras en la canción. Los hallazgos se interpretan en relación con las teorías del aprendizaje de palabras y cómo las diferencias en las características entre el habla y la canción pueden relacionarse con la variabilidad en la eficacia al adquirir nuevas palabras.

Palabras clave: aprendizaje de palabras; canciones; alumnos de educación preescolar

1. Introduction

In many cultures around the world, young children are frequently exposed to songs; this includes during the preschool years when children are preparing to start school. Often, these songs are meant to help teach new information (e.g., the letters in the alphabet, the colors of the rainbow, etc.). While preschoolers appear to enjoy songs, there is little empirical data supporting whether presenting new concepts (and specifically words) in melodies, as opposed to spoken sentences, leads to better learning.

Speech and song are features of human communication (e.g., Brandt et al., 2012; Trehub, 2019). The similarity between speaking and singing is noteworthy; both require the same structural components (e.g., vocal folds, respiration), and have the ability to express feeling and emotion through shared acoustic properties (i.e., prosody, pitch) (Juslin & Laukka, 2003; Poeppel & Assaneo, 2020; Quinto et al., 2013; Thompson et al., 2004). The ability to process speech and song has been proposed to rely on a shared set of cognitive and neural abilities (e.g., Fedorenko et al., 2009; Fiveash et al., 2021; Thompson et al., 2012). Further, exposure to songs has been suggested to lead to advantages in neural processing that are related to language development (Brandt et al., 2012; Dehaene-Lambertz et al., 2010; Zhao & Kuhl, 2016). Moreover, song has been proposed to play a role in social and emotional development (see Smith & Kong, 2024, for review). Thus, speech and song are tools used to convey meaning. It is perhaps for this reason that there has been a rise in popularity of word learning through song programs (Debreceny, 2015; Governor et al., 2013; gymbobuzz, 2021), and a consistent use of song in school curricula (Kirby et al., 2023; Rajan, 2017).

The idea that song may benefit learning likely stems from research that has shown that children prefer, and learn more, from speech registers that are more "musical," such as child-directed speech (CDS). CDS is characterized by a slower rate of speech, higher pitch, elongation of vowels, and its increased rhythmicity compared to adult-directed speech (ADS) (Fernald & Simon, 1984; Stern et al., 1982), making it appear more melodic or song-like (Fernald et al., 1992). Young children show a preference for CDS compared to ADS (Frank et al., 2020), and research has suggested that CDS facilitates word learning in comparison to ADS (Ma et al., 2011; Singh et al., 2009; Thiessen et al., 2005). Very often, songs, particularly child-directed songs, share similarities with CDS, such as a slower tempo, higher pitch, and repetition (Trainor et al., 1997; Trehub et al., 1997). These shared characteristics would suggest that children's songs, like CDS, might facilitate vocabulary learning.

Studies exploring learning in song in the first language have primarily been conducted with infants or older school-age children. While word learning in song likely takes place in classrooms of both first and second language learning, we focus on first language development in the present study and subsequent literature review. However, it should be noted that prior work has found little difference between different types of rhythmic input on the depth of word learning between first and second language learners (e.g., Lawson-Adams et al., 2022).

One study with 6.5–8-month-old infants found that infants were able to recognize a familiar number sequence when it was sung, but not when it was spoken (Thiessen & Saffran, 2009). Another study with 11-month-olds tested infants' ability to detect changes in a sequence of notes or speech sounds (Lebedeva & Kuhl, 2010). This study found a facilitatory effect of song, with changes in the note sequence being detected but not changes to the speech sequence. These findings are supported by physiological evidence, which used event-related potentials (ERPs) to examine word segmentation in 10-month-old infants

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

(Snijders et al., 2020). This paper found that when familiarized with speech versus song, infants demonstrated similar levels of recognition for words in both conditions. Though these findings demonstrate some effect of song on sequence learning and segmentation, they have been conducted with infants and have not directly assessed the creation of a word-item link.

Studies with school-age children have produced conflicting data. On the one hand, there appears to be a facilitatory effect of songs compared to speech during word learning (Chou, 2014; Davis & Fan, 2016; Good et al., 2015; Zhou & Li, 2017). However, other studies report a null effect of songs during word learning compared to speech alone (Albaladejo et al., 2018; Heidari & Araghi, 2015; Leśniewska & Pichette, 2016). Thus, the role of song during word learning is unclear for school-age children.

To our knowledge, only two studies have explored word-learning in song in a population of toddlers and preschool-age children. One study explored word learning through the presentation of novel objects in children from 1 to 5 years old. The novel items were presented and named either in song or in ADS (Ma et al., 2023). The authors found that the novel words were learned better in song than in ADS, and that the words learned in song were retained at a delayed testing measure, whereas words learned in ADS were not. However, this work did not provide information as to whether song necessarily leads to better learning compared to other types of speech, particularly CDS. This is important, given that young children are mostly spoken to in CDS rather than in ADS.

We explored this question in a prior study (Morini & Blair, 2021), in which children were taught novel word-object relations in song and in spoken sentences (using CDS prosody). This study utilized a Preferential Looking Paradigm (PLP) and was conducted with toddlers between the ages of 29–32 months and preschoolers between 47 and 50 months. Both groups of learners demonstrated the ability to learn novel words in both song and CDS. Critically, the 29–32-month-olds showed no statistical difference in learning between speech and song, whereas the 47–50-month-olds learned the novel words better in the spoken condition compared to the song condition. One important limitation of this prior work was that the tune that was used for the song condition ("Old MacDonald Had a Farm") had minimal pitch contrast when the target words were introduced in the song, while the spoken (CDS) condition had prosodic variation that emphasized the target word within the sentence. This prosodic variation may have facilitated the ability to segment and attend to the novel word (Song et al., 2010). Additionally, presenting children with a familiar melody that was matched with an unexpected lyric change may have made the task more difficult due to the stimuli not aligning with the child's expectation. In the present study, we address these limitations by (i) using an unfamiliar melody, and (ii) controlling for pitch variation between the spoken and song stimuli.

2. Experiment

We examined the role of song during the acquisition of novel word-object relations in preschool-age children using a within-subjects design. Specifically, we wanted to understand whether controlling for pitch differences would lead to better or equivalent word learning in song compared to CDS in preschool-age children. We chose to only run the 47–50-month-olds, as in the prior study (Morini & Blair, 2021), this age group displayed learning differences between the song and CDS conditions, whereas the younger children did not. The design of the present study was identical to that of Morini and Blair (2021),

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

with the key difference being the melody used to teach the novel words in the song condition.

3. Methods

3.1. Participants

A total of 32 children (14 female, 17 male, 1 other), between the ages of 47–50 months (M = 48.6 SD = .85) participated. A sensitivity analysis revealed that this sample is large enough to detect differences of a medium effect size as found in the Morini and Blair (2021) paper. Of the participants, 22 were White, 3 were Black/African American, 1 was Hispanic/Latino, 2 were Asian, 3 were Mixed Race, and 1 declined to answer. Socio-economic status was determined via maternal education, and on average, mothers had 18 years of education (SD = 1.77), the equivalent of a master's degree. Data from an additional 28 children were dropped due to technical issues (N = 12), child fussiness/inattention (N = 13), side bias (N = 1) and ineligibility for the study that was not discovered until the time of study running (N = 2). This dropout rate is not unusual for work using the PLP (Schmale et al., 2012), and matches that of prior work Morini and Blair (2021). Moreover, there were no differences in dropout rate based on condition (i.e., speech versus song). Children were raised in monolingual English-speaking homes and did not have any diagnosed disabilities or language disorders based on parental report. Children completed the study virtually from their own home via a synchronous Zoom appointment and needed to have access to a computer with at least a 12-inch screen and a webcam, and a reliable internet connection.

3.2. Stimuli

Four videos of novel objects were used as the visual stimuli. In the videos, the objects were slowly moved from side to side to maintain visual attention. The four objects were paired up, all objects were made of the same material (i.e., wood), were similar sizes, and had equivalent anticipated salience. Each object was notably different from the others, as they were all different solid colours.

The auditory stimuli were recorded by a female native speaker of American English. Training sentences were either spoken using CDS or sung to a tune that matched the pitch of the CDS sentence (see Figure 1).

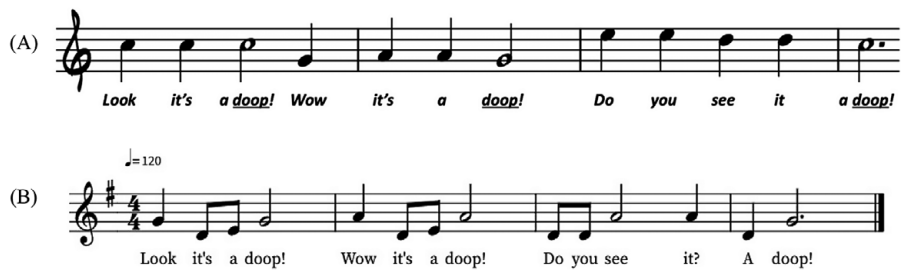

Figure 1. The melody used in the prior study (Morini & Blair, 2021) (A), and in the current study (B).

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

The training sentences were comprised of the carrier phrase ("Look! It's a ___. Wow it's a _____. Do you see it? A _____."). The target word was embedded into the carrier phrase. The four novel target words were one syllable in length and were created following English phonotactic rules (i.e., "doop," "neff," "shoon," "fim"). The carrier phrase in testing sentences provided a directive for children to look at one of the two items on the screen. ("Look at the _____! Do you see the _____? Where is that _____? _____!"). All test phrases were produced in spoken sentences using CDS prosody. In both testing and training trials, the onset of the first instance of the target word was 1.4 seconds after the onset of speech. All trials had a duration of 7.5 seconds. For more information, see Morini and Blair (2021).

3.3. Procedure

The study paradigm was comprised of four testing blocks, two in the song condition and two in the spoken (CDS) training condition. Blocks 1 and 2 taught and tested a new word: one in the song condition, and one in the CDS condition. Blocks 3 and 4 were repetitions of the first two blocks (i.e., Block 1 = Block 3, Block 2 = Block 4). Each block began with a silent trial where an object pair was shown on the screen to examine any object or side biases. Next, three identical training trials were presented. In training trials, a single item was presented on the screen and was accompanied by a training sentence either in song or CDS (Figure 2). Testing took place immediately after training trials. Children were tested on the trained item and on the second untrained item, which was given an unfamiliar name. These untrained trials were included, as if children learned the trained item, they should look longer at the new item when a novel name was presented, due to the principle of mutual exclusivity (Markman & Wachtel, 1988). The assumption of mutual exclusivity is that children assume that each item has only one label and will assign a novel label to the untrained item. Both the trained and untrained test trials assessed learning of the trained word-object pair via direct recall and mutual exclusivity testing, respectively. Additionally, this approach controls for trained object preferences that may arise due to repeated familiarization of the trained item. In between all trial types, an 8-second attention getter with a black background was presented (Figure 2).

For participants to be included in the final sample, children needed to have had complete, analysable data for at least one block in each of the training conditions. We counterbalanced for the following parameters across participants (i) which words were trained versus untrained words, (ii) whether trained or untrained testing trials were presented first, (iii) whether trials in the song condition were presented in blocks 1 and 3 or blocks 2 and 4, (iv) which object was assigned which label, and (v) the position of the items on the screen (i.e., if the items were presented on the left or right at testing). See Tables S1 and S2 in Supplemental Materials for confirmation of null effects of the counterbalanced items on the study results.

During the study appointment, caregivers were asked to find a quiet room and eliminate any possible distractors during the appointment (i.e., turning off the TV/music). All experimenters followed a written testing protocol. The appointment began with a check of the lighting, the webcam, and the audio of the participating family's computer, utilizing a 30-second video of a spinning whale, which was sent to the family in the Zoom chat. In this video, music played, and the background changed from black to white. This background change allowed the experimenter to note if changes in brightness were visible on the child's face when the screen went from black to white from the webcam. This video

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

Trial #	Audio	Visual
1	Silence	
2,3,4	"Look it's a doop! Wow, it's a doop. Do you see it? A doop!"	
5	"Look at the doop! Do you see the doop? Where is that doop? Doop!"	
6	"Look at the neff! Do you see the neff? Where is that neff? Neff!"	

Figure 2. An example study block.

was additionally used to test the audio level, as the music was at the same intensity level as the stimuli in the word learning task. Caregivers were asked to adjust their computer's volume so that the music could be heard at a comfortable listening level.

After completing the checks, experimenters turned off their cameras and muted themselves not to be seen or heard during the study appointment. A link to the study video, which contained all training and testing trials as well as attention getters, was sent to the parents through the chat in Zoom. Parents were instructed to record the session locally on their computer using pre-installed software (e.g., Quicktime for Mac and Camera for Windows). After beginning their recording, caregivers played the study video and closed their eyes for the duration of the video. The children completed the task seated on their caregiver's lap. Once the video had finished playing, the experimenters turned their audio and camera's back on, and guided the caregivers through uploading their video of the testing session through a secure link.

3.4. Data coding

All participant videos were coded offline, frame by frame, by two trained coders utilizing the Datavyu coding software (Datavyu Team, 2014). All coded files were checked for

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

coder reliability, and any trials where there were discrepancies of over 0.5 seconds were recoded by a trained third coder, which is common practice for hand-coded data (e.g., Newman et al., 2018). In this study, 16.1% of trials required a third coder, which is not uncommon in data coded from young children (Newman et al., 2018), or online studies (Morini & Blair, 2021).

4. Results

We began by looking for outliers in the data using a z-score method, in which values that were over or under 2 standard deviations from the mean were dropped (e.g., Venkataanusha et al., 2019). After removing the 8 outlier data points, the remaining data included 244 data points from 32 children. Of these data points, 122 of these observations were from the song condition, and the other 122 were from the spoken condition.

Next, we examined the children's looking patterns during the baseline trials to ensure that there were no pre-existing side or item biases. We found that children looked to the item on the left side of the screen 49% of the time (SD = .11) and to the item on the right side 51% of the time (SD = .11). As children were given no looking instructions during baseline trials, these looking patterns suggest that there were no overall side biases. We then calculated accuracy based on the proportion of time the participants looked toward the target object compared to the competitor item during testing (i.e., accuracy), over a time window of 367–5100 ms after the onset of the target word, across all trials of the same condition (Morini & Blair, 2021). Target items varied across testing conditions and included the trained object and the untrained object, depending on the type of testing trial and the item that was requested. Each of these objects was the "correct" object on one of the two test trials, and if the children adequately learned the trained word, they should be able to look toward the target item in both test trials (e.g., Markman & Wachtel, 1988).

Further exploration of the data utilized two-tailed single-sample t-tests comparing child performance in speech and in song to chance (in this case, 50%), to determine whether children had been able to learn the novel words in each condition. These t-tests demonstrated that both training in speech ($t(31) = 8.36$, Cohen's $d = 1.48$, $p < .001$) and in song ($t(31) = 5.58$, Cohen's $d = .99$, $p < .001$) led to learning above chance. This indicated that children were able to successfully learn the novel words in both training conditions. While children demonstrated learning in both conditions, on average, children looked longer at the target item for words trained in IDS (M = .69, SD = .20), compared to song (M = .63, SD = .22) (Figure 3). For further visualization of the data, see Figure 1 in Supplemental Materials.

To examine the role of song during novel word learning, we ran a mixed-effects model using the lme4 package in R (Bates et al., 2015). The mixed-effects model compared accuracy across both learning conditions (speech versus song). The type of test trial (trained versus untrained), and whether the testing block (i.e., the first instance versus the second instance of learning in song/speech) were included in the model as interactions (formula = Accuracy~Condition*(Training+Block) + (Condition|ID)). All factor items were deviation coded within the model. Within the model, the following factors were coded as 0: Training in Song, Testing in Block 1, Untrained Testing, and these factors were coded as 1: Training in Speech, Testing in Block 2, Trained Testing. Additionally, random slopes were included for condition by participant to take into account individual differences related to both the condition of learning and the child. A complete random effects structure could not be run due to convergence issues, which is further acknowledged in the discussion.

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

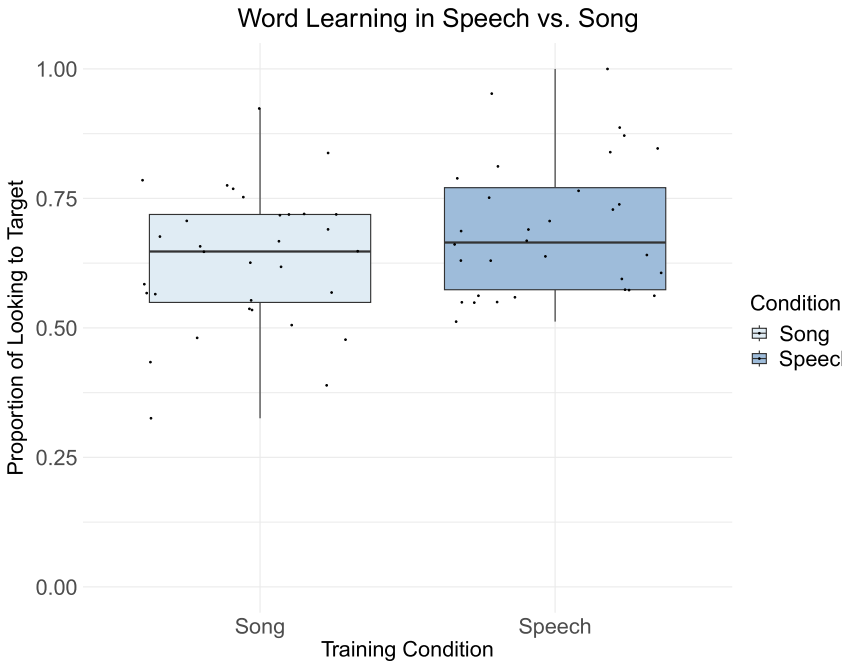

Figure 3. A graph of accuracy in word learning in speech and in song.

The mixed effects model showed a significant main effect of training condition, with child performance in the spoken condition leading to higher accuracy at test than when the novel word was learned in the song condition ($\beta = .11$, $\eta_p^2 = .14$, $p < .05$). No other comparisons were significant, suggesting that the type of testing condition (i.e., trained versus untrained words) and the block did not significantly affect performance on the task (see Table 1). Further, a comparison of the data from the present study and the 2021 paper demonstrated no significant differences in performance between the two studies by group ($\beta = .00$, $\eta_p^2 = .00$, $p > .05$)., or group by condition interactions ($\beta = .00$, $\eta_p^2 = .00$, $p > .05$). This suggests that even with pitch changes, learning in song was essentially the same between song conditions in the present and past data sets (see Table S3 in Supplemental Materials).

5. Discussion

This study expanded on prior work examining the role of song on word learning in preschool-age children. Furthermore, we evaluated whether prior findings suggesting that preschoolers learned new words better in speech than in song were due to pitch differences in the stimuli. The present work indicated that even when matching pitch, preschoolers demonstrated better learning of novel words in speech compared to song, as seen in the initial study (Morini & Blair, 2021). This suggests that differences in learning between speech and song are found in the Morini and Blair (2021) study were not caused by an enhanced identification of the novel word due to the variation in pitch between the spoken and song conditions. This aligns with prior findings which suggest that prosodic

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

Table 1. Model of performance in speech versus song (bolded items are statistically significant p < .05)

Predictors	Estimates	CI	*p*
(Intercept)	0.62	0.56–0.69	**<0.001**
Condition [Speech]	0.11	0.02–0.19	**0.015**
Block [Two]	0.04	−0.03–0.11	0.246
Training [Trained]	0.00	−0.07–0.07	0.968
Condition [Speech] × Block [Two]	−0.02	−0.12–0.07	0.616
Condition [Speech] × Training [Trained]	−0.07	−0.17–0.02	0.143
Random effects			
σ^2		0.04	
$\tau_{00\ ID}$		0.01	
$\tau_{11\ ID.ConditionSpeech}$		0.00	
$\rho_{01\ ID}$		−0.51	
ICC		0.15	
N_{ID}		32	
Observations		238	
Marginal R^2/Conditional R^2		0.040/0.179	

cues, while important for speech segmentation, appear to be most influential for infants under 12 months old (Männel & Friederici, 2013). Further, Snijders et al. (2020) posed that pitch changes were not critical for segmenting familiar words from spoken sentences, nor from song. Thus, other factors may be leading to the differences seen in learning between speech and song.

One possibility is that the variability seen between learning in speech versus song may be related to theories of deep learning (e.g., Calvert, 2001). Calvert (2001) proposed that learning takes place in levels. Some of these levels are more superficial, such as verbatim memory; however, other levels of learning are deeper, such as the ability to encode and retrieve information about prior events. For vocabulary learning, superficial learning may lead to recalling the sound sequence, but not necessarily the word-object link. Deeper learning, in contrast, would be not only creating a word-object link, but also encoding greater meaning about that object. While a song may appear to be "catchy," the listener is not only hearing the content of the utterance, but now a melody and a cadence that differ from typical CDS. Because of this, during song, children may not as easily attend to the content, or the words being sung, as they are also attending to other qualities of the song. Future work could explore whether children are attending to and learning about other properties of song (e.g., the melody) more than the novel word being presented by teaching the names of novel objects with different melodies. At the test, they could be instructed to identify the previously taught item with the same melody used at training, or with a new melody that was not present at training. If children attach the melody more to the object rather than the name, their performance in the same melody condition would be expected to be higher than in the different melody testing condition. It should be noted, however, that in the present study, both song and speech conditions led to learning above

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

chance. Therefore, while learning in song was more difficult than in speech, it did not prevent learning.

Another possibility is that children may have had difficulty generalizing the information learned in the song condition, as all testing took place in CDS. The rationale for this methodology was that even if children learned new words in song, they need to be able to then successfully recognize the word in other speech contexts. While it is unlikely, given that it has been shown that even infants are able to recognize words across different registers (e.g., happy voices versus neutral voices) (e.g., Singh, 2008), it is possible that the mismatch between training and testing may have made testing in the song condition more difficult than in the speech condition. Future work should explore novel word learning in speech and song with testing conditions that match the conditions of the training.

Additionally, there are some limitations that are worth discussing. First, all our participants were within a tight age range. Older children or younger children may respond differently to speech and song for word learning, and therefore, these populations should also be tested in future work. Additionally, our population was not very diverse in race, socioeconomic status, or in their language background (all were monolingual). Future work in this space should aim to work with more racially, ethnically, and linguistically diverse populations, as well as with participants from varied socioeconomic backgrounds. Moreover, our sample size and number of items did not allow for a complete random structure to be included in our analyses, meaning that we were unable to account for random slopes due to the within participant factors of testing Block, and Trained versus Untrained testing. Additional studies should be conducted with larger sample sizes and a wider range of items tested.

Second, the methodology of the study could be made more naturalistic. In our design, children rapidly learned the name of an item over 3 identical training trials and then immediately went into testing. This is not often how children learn the names of items in a real-world context, where they instead may learn the name of an item slowly over days and then have no formal testing. Future studies should explore how children are able to learn the novel names of items in speech and in song in scenarios that are more ecologically valid.

Third, our data only measures immediate learning and does not examine retention effects. It is possible that consolidation of the word-object relations may be affected by training in the spoken or song conditions, as prior work has found differences in the retention of newly learned words when they were taught in song or in speech (ADS) (Ma et al., 2023). Therefore, more work is needed to determine if there are retention effects of learning in song versus in spoken (CDS) language.

Finally, our study had a high drop-out rate, with data from 28 children being lost due to technical issues or fussiness. While not necessarily atypical in work with preschool-age children, this high drop-out rate may affect the generalizability of our findings, as many children were unable to provide sufficient data for the study. Future work should aim for less data loss, perhaps by modifying the design of the study.

6. Conclusion

The present work explored whether matching pitch and prosodic cues would affect how well preschool-age children were able to learn novel word-object relations in speech and in song. It was found that children learned the novel words better when they were trained in speech rather than in song, which aligns with prior work. These findings suggest that when aiming to teach children new words, using song may not be the most beneficial tool, compared to CDS (at least not for initial learning). This is increasingly relevant with the

https://doi.org/10.1017/S0305000925100081 Published online by Cambridge University Press

rise of "learning through song" programs, and can inform parent and teacher practices when approaching vocabulary learning. More research is still needed in this space and should examine learning in speech and song with different populations, different testing measures, and more naturalistic training. Further, future work should examine the role of learning in song on retention of newly learned word-object relations, rather than only immediate testing.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0305000925100081.

Acknowledgements. We are grateful for the support of the entire Speech Language Acquisition and Multilingualism Lab and their help with contacting families, running study appointments, and coding data. Particularly, we want to thank Meli R. Ayala and Karla Marie Mercedes for their help scheduling participants.

Competing interests. The authors declare none.

References

Albaladejo, S., **Coyle, Y.**, & **De Larios, J. R.** (2018). Songs, stories, and vocabulary acquisition in preschool learners of English as a foreign language. *System*, **76**, 116–128. https://doi.org/10.1016/j.system.2018.05.002.

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

Brandt, A., **Gebrian, M.**, & **Slevc, L. R.** (2012). Music and early language acquisition. *Frontiers in Psychology*, **3**, 1–17. https://doi.org/10.3389/fpsyg.2012.00327.

Calvert, S. L. (2001). Impact of televised songs on children's and young adults' memory of educational content. *Media Psychology*, **3**(4), 4. https://doi.org/10.1207/S1532785XMEP0304_02.

Chou, M. (2014). Assessing English vocabulary and enhancing young English as a foreign language (EFL) learners' motivation through games, songs, and stories. *Education 3–13*, **42**(3), 284–297. https://doi.org/10.1080/03004279.2012.680899.

Datavyu Team. (2014). Datavyu: A video coding tool. In *Databrary project*. New York University. www.datavyu.org

Davis, G. M., & **Fan, W.** (2016). English vocabulary acquisition through songs in Chinese kindergarten students. *Chinese Journal of Applied Linguistics*, **39**(1), 59–71. https://doi.org/10.1515/cjal-2016-0004.

Debreceny, A. (2015). Song as saga: Curriculum-based songs for learning. *2nd International Conference on Education and Social Sciences*, International Organization Center of Academic Research, Istanbul (pp. 301–310).

Dehaene-Lambertz, G., **Montavont, A.**, **Jobert, A.**, **Allirol, L.**, **Dubois, J.**, **Hertz-Pannier, L.**, & **Dehaene, S.** (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, **114**(2), 53–65. https://doi.org/10.1016/j.bandl.2009.09.003.

Fedorenko, E., **Patel, A.**, **Casasanto, D.**, **Winawer, J.**, & **Gibson, E.** (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, **37**(1), 1–9. https://doi.org/10.3758/MC.37.1.1.

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

Fernald, A., & **Simon, T.** (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, **20**(1), 104–113.

Fiveash, A., **Bedoin, N.**, **Gordon, R. L.**, & **Tillmann, B.** (2021). Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology*, **35**(8), 771–791. https://doi.org/10.1037/neu0000766.

Frank, M. C., **Alcock, K. J.**, **Arias-Trejo, N.**, **Aschersleben, G.**, **Baldwin, D.**, **Barbu, S.**, **Bergelson, E.**, **Bergmann, C.**, **Black, A. K.**, **Blything, R.**, **Böhland, M. P.**, **Bolitho, P.**, **Borovsky, A.**, **Brady, S. M.**, **Braun, B.**, **Brown, A.**, **Byers-Heinlein, K.**, **Campbell, L. E.**, **Cashon, C.**, & **Davies, C.** (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, **3**(1), 24–52.

Good, A. J., **Russo, F. A.**, & **Sullivan, J.** (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, **43**(5), 627–640.

Governor, D., **Hall, J.**, & **Jackson, D.** (2013). Teaching and learning science through Song: Exploring the experiences of students and teachers. *International Journal of Science Education*, **35**(18), 3117–3140. https://doi.org/10.1080/09500693.2012.690542.

gymbobuzz. (2021). 45 Years of Gymboree Play & Music. *Gymbobuzz*. https://gymbobuzz.gymboreeclasses.com/2021/05/27/45-years-of-gymboree-play-music/

Heidari, A., & **Araghi, S. M.** (2015). A comparative study of the effects of songs and pictures on Iranian EFL learners' L2 vocabulary acquisition. *Journal of Applied Linguistics and Language Research*, **2**(7), 24–35.

Juslin, P. N., & **Laukka, P.** (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, **129**(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770.

Kirby, A. L., **Dahbi, M.**, **Surrain, S.**, **Rowe, M. L.**, & **Luk, G.** (2023). Music uses in preschool classrooms in the U.S.: A multiple-methods study. *Early Childhood Education Journal*, **51**(3), 515–529. https://doi.org/10.1007/s10643-022-01309-2.

Lawson-Adams, J., **Dickenson, D. K.**, & **Donner, J. K.** (2022). Sing it or speak it?: The effects of sung and rhythmically spoken songs on preschool children's word learning. *Early Childhood Research Quarterly*, **58**(1), 87–102. https://doi.org/10.1016/j.ecresq.2021.06.008.

Lebedeva, G. C., & **Kuhl, P. K.** (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, **33**(4), 4. https://doi.org/10.1016/j.infbeh.2010.04.006.

Leśniewska, J., & **Pichette, F.** (2016). Songs vs. stories: Impact of input sources on ESL vocabulary acquisition by preliterate children. *International Journal of Bilingual Education and Bilingualism*, **19**(1), 18–34. https://doi.org/10.1080/13670050.2014.960360.

Ma, W., **Bowers, L.**, **Behrend, D.**, **Hellmuth Margulis, E.**, & **Forde Thompson, W.** (2023). Child word learning in song and speech. *Quarterly Journal of Experimental Psychology*, **77**(2), 343–362. https://doi.org/10.1177/17470218231172494

Ma, W., **Golinkoff, R. M.**, **Houston, D. M.**, & **Hirsh-Pasek, K.** (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, **7**(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Männel, C., & **Friederici, A. D.** (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, **49**(10), 2788–2798. https://doi.org/10.1016/j.cortex.2013.09.003.

Markman, E. M., & **Wachtel, G. F.** (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, **20**(2), 2. https://doi.org/10.1016/0010-0285(88)90017-5.

Morini, G., & **Blair, M.** (2021). Webcams, songs, and vocabulary learning: A comparison of in-person and remote data collection as a way of moving forward with child-language research. *Frontiers in Psychology*, **12**, 702819. https://doi.org/10.3389/fpsyg.2021.702819.

Newman, R. S., **Morini, G.**, **Kozlovsky, P.**, & **Panza, S.** (2018). Foreign accent and toddlers' word learning: The effect of phonological contrast. *Language Learning and Development*, **14**(2), 2. https://doi.org/10.1080/15475441.2017.1412831.

Poeppel, D., & **Assaneo, M. F.** (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, **21**(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4.

Quinto, L., **Thompson, W. F.**, & **Keating, F. L.** (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, **4**, 1–8. https://doi.org/10.3389/fpsyg.2013.00184.

Rajan, R. S. (2017). Preschool teachers' use of music in the classroom: A survey of Park District preschool programs. *Journal of Music Teacher Education*, **27**(1), 89–102. https://doi.org/10.1177/1057083717716687.

Schmale, R., **Cristia, A.**, & **Seidl, A.** (2012). Toddlers recognize words in an unfamiliar accent after brief exposure: Brief exposure to an unfamiliar accent. *Developmental Science*, **15**(6), 732–738. https://doi.org/10.1111/j.1467-7687.2012.01175.x

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, **106**(2), 833–870. https://doi.org/10.1016/j.cognition.2007.05.002

Singh, L., **Nestor, S.**, **Parikh, C.**, & **Yull, A.** (2009). Influences of infant-directed speech on early word recognition. *Infancy*, **14**, 654–666. https://doi.org/10.1080/15250000903263973.

Smith, A. R., & **Kong, K. L.** (2024). Music enrichment programs may promote early language development by enhancing parent responsiveness: A narrative review. *Child Development Perspectives*, **19**(1), 20–29. https://doi.org/10.1111/cdep.12519

Snijders, T. M., **Benders, T.**, & **Fikkert, P.** (2020). Infants segment words from songs—An EEG study. *Brain Sciences*, **10**(1), 39. https://doi.org/10.3390/brainsci10010039

Song, J. Y., **Demuth, K.**, & **Morgan, J.** (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, **128**(1), 389–400. https://doi.org/10.1121/1.3419786.

Stern, D. N., **Spieker, S.**, & **MacKain, K.** (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, **18**(5), 727–735.

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.

Thiessen, E. D., & **Saffran, J. R.** (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, **1169**(1), 225–233. https://doi.org/10.1111/j.1749-6632.2009.04547.x

Thompson, W. F., **Marin, M. M.**, & **Stewart, L.** (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, **109**(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109.

Thompson, W. F., **Schellenberg, E. G.**, & **Husain, G.** (2004). Decoding speech prosody: Do music lessons help? *Emotion*, **4**(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46.

Trainor, L. J., **Clark, E. D.**, **Huntley, A.**, & **Adams, B. A.** (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, **20**(3), 383–396. https://doi.org/10.1016/S0163-6383(97)90009-6.

Trehub, S. E. (2019). Nurturing infants with music. *International Journal of Music in Early Childhood*, **14**(1), 9–15). https://doi.org/10.1386/ijmec.14.1.9_1

Trehub, S. E., **Unyk, A. M.**, **Kamenetsky, S. B.**, **Hill, D. S.**, **Trainor, L. J.**, **Henderson, J. L.**, & **Saraza, M.** (1997). Mothers' and fathers' singing to infants. *Developmental Psychology*, **33**(3), 500–507.

Venkataanusha, P., **Anuradha, C.**, **Chandra Murty, P. S. R.**, & **Chebrolu, S. K.** (2019). Detecting outliers in high dimensional data sets using z-score methodology. *International Journal of Innovative Technology and Exploring Engineering*, **9**(1), 48–53. https://doi.org/10.35940/ijitee.A3910.119119.

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*, **113**(19), 5212–5217. https://doi.org/10.1073/pnas.1603984113.

Zhou, W., & **Li, G.** (2017). The effects of shared singing picture book instruction on chinese immersion kindergarteners' spoken vocabulary recall and retention. *Frontiers in Education*, **12**(1), 29–51. https://doi.org/10.3868/s110-006-017-0003-5.

Cite this article: Blair, M., Hawtof, L., & Morini, G. (2025). Spoken or sung? Examining word learning in child-directed speech and in song. *Journal of Child Language* 1–13, https://doi.org/10.1017/S0305000925100081