The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception

Sarah C. Creel¹

Published online: 11 January 2019
© The Psychonomic Society, Inc. 2019

Abstract

How do learners build up auditory pattern knowledge? Findings from children's spoken word learning suggest more robust auditory representations for highly familiar words than for newly learned words. This argues against spoken language learning as a process of simply acquiring a fixed set of speech sound categories, suggesting instead that specific words may be the relevant units. More generally, one might state this as the *specific-learning hypothesis*—that acquiring sound pattern knowledge involves learning specific patterns, rather than abstract pattern components. To understand the nature of human language knowledge, it is important to determine whether this specific learning reflects processes unique to spoken language learning or instead reflects more general auditory-learning processes. To investigate whether the specific-learning hypothesis extends to auditory pattern learning more generally, the present study tested the perceptual processing of familiar melodies versus carefully matched unfamiliar melodies. Children performed better at both audiovisual mapping (Exp. 1) and same–different auditory discrimination (Exp. 2) when hearing familiar melodies than when hearing unfamiliar melodies. This is consistent with the specific-learning hypothesis and with exemplar-style general-auditory accounts of pattern learning, although alternative explanations are possible.

Keywords Music cognition · Sound recognition · Speech perception · Spoken word recognition

A major question in both speech processing and spoken language development is the extent to which the auditory representations of language depend on neurally prespecified sound patterns restricted to the domain of language, or instead represent the outcome of more general learning principles. Diehl and colleagues (Diehl & Walsh, 1989; Hay & Diehl, 2007; Kluender, Diehl, & Wright, 1988) have written extensively about the overlap between speech processing and general auditory processes. Their research suggests that speech sound patterns depend on or emerge from general auditory learning processes, and that speech sound patterns do not depend on having prespecified categories (see especially Kluender, Diehl, & Killeen, 1987, who showed speech sound learning in quail). According to this second account, one should see

parallels between auditory pattern learning in speech and in nonspeech.

In this spirit, the present work explores the learning of sound patterns of nonspeech auditory stimuli. To further substantiate an account of spoken language learning in terms of general auditory processes, we ask whether nonspeech auditory learning shows analogous pattern-learning processes. We explore an *exemplar view* of the learning process. According to exemplar accounts (e.g., Goldinger, 1998), representations emerge from accruing many specific exemplars, or large numbers of neural traces. Only after collecting a large number of representations do broader-scale patterns—such as speech sounds apart from particular word contexts, or the generic properties of musical scales—emerge. This suggests that early in the learning process—for example, in childhood—learners may distinguish *well-known* sound patterns that differ subtly, while failing to distinguish *novel patterns* that differ by the same characteristics.

✉ Sarah C. Creel
screel@ucsd.edu

¹ Cognitive Science, University of California, San Diego, 9500
Gilman Drive Mail Code 0515, La Jolla, CA 92093, USA

It is worth contrasting exemplar accounts with other explanatory frameworks, such as prototype theories of memory formation. In an exemplar account, recognition is computed by a composite of traces that are each activated in proportion to their similarity to the input. In a prototype account, recognition is accomplished by comparing a new instance to each category's central tendency. Thus, both types of accounts are probabilistic. One area in which they differ, though, is that many prototype or prototype-like accounts of pattern learning specify a low-level unit of analysis in terms of speech sounds (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), or in music, in terms of such properties as musical key membership (e.g., Trainor & Trehub, 1994). By contrast, exemplar accounts often take events, such as entire words, as the relevant unit of analysis (see, e.g., Goldinger, 1998). A prototype account that posits melodies as the unit of analysis might make similar predictions in the present study. Still, exemplar accounts also have other appealing properties, in that the perceptual category does not need to be preassigned but can emerge, unsupervised, from the input, and that category variability is implicitly preserved. However, in the present case the appeal of exemplar approaches is less their exact similarity computations than their basic unit of analysis.

Empirical support for exemplar-like auditory event memory

One set of findings that has suggested that auditory event memories consist of specific exemplars comes from the domain of early child word learning. Infants (14-month-olds) can distinguish similar speech sounds such as /b/ and /d/ in immediate discrimination (detecting a change to “dih” after many repetitions of “bih”), but they perform poorly at learning labels distinguished by these words (*bih* vs. *dih*) at 14 months, only succeeding at 17–20 months (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). One interpretation of this disconnect between discrimination and word learning is that word learning requires maintaining longer-term sound-pattern representations than does discrimination, which requires only short-term representations. However, in contrast with the novel-word-learning findings, *familiar* words that are distinguished by single speech sounds (e.g., *ball* vs. *doll*) are readily recognized by children around 14 months (Fennell & Werker, 2003; Swingley & Aslin, 2002). Note that the critical contrast here is between two similar studies, rather than within a single study. On the one hand, Stager and Werker presented children with two unfamiliar pictures, each with an unfamiliar label—*bih* or *dih*—and then tested their learning of the labels. Fennell and Werker used exactly the same procedure, and exactly the same b–d speech-sound contrast, but used instead the familiar words *ball* and *doll* and pictures of a ball and a doll. Whereas the *bih/dih*

children failed at 14 months (Stager & Werker, 1997), the *ball/doll* children succeeded (Fennell & Werker, 2003), indicating that they could use the b–d contrast to tell words apart, but only when it was embedded in familiar words. This suggests that children may not yet have separable representations of individual speech sounds, but instead have representations of entire words, consistent with an exemplar account.

Music perception and specific auditory memory

There is evidence of a similar phenomenon in the music perception literature. For example, McFadden and Callaway (1999) found, in a series of psychophysical experiments, that adult listeners were more sensitive to subtle changes in familiar musical (and speech) materials than in matched unfamiliar materials. The evidence from child music perception is suggestive but not conclusive. Research by Trainor and colleagues suggests that 4-year-olds are less perceptually sensitive than adults to musical structure violations in unfamiliar melodies (Trainor & Trehub, 1994), but are excellent at detecting musical structure violations in familiar melodies, specifically “Twinkle Twinkle Little Star” (Corrigall & Trainor, 2010). In a related finding that mirrors the Stager and Werker (1997) result of discrimination ability without word-learning ability, Creel (2014c, 2016) reported that preschool-aged children can auditorily discriminate some musical patterns (rising vs. falling pitches) that they cannot associate with visual objects (cartoons; Creel, 2014c, 2016; see also Pajak, Creel, & Levy, 2016, for a similar effect in adult second-language word learning). This is especially striking in that the hard-to-associate sounds differ in *pitch contour*, a musical property to which even young infants are sensitive (Trehub, Bull, & Thorpe, 1984),¹ and one regarded as central to melodic identity (Dowling, 1978). Of course, Creel's (2014c, 2016) audio-visual association tasks used *unfamiliar* brief melodies, raising the possibility that more familiar melodies would be more associable.

These musical and linguistic findings seem at odds with accounts of learning as a process of acquiring speech sound categories (Werker & Tees, 1984) or acquiring abstract musical knowledge such as scale structure (Lynch, Eilers, Oller, & Urbano, 1990). They are easier to square with an account that children accrue specific exemplars of auditory patterns, such as melodies and words, rather than directly learning more abstract structures (major scales or phonology). Thus, the relatively weaker performance with less-familiar materials is

¹ Creel (2014c, 2016; see also Creel & Quam, 2015) have attributed this seeming developmental disconnect—good performance by infants but some difficulties in preschoolers—to substantial differences in the infant versus older-child test paradigms. One cannot perform a habituation test with a 4-year-old, nor can one conduct a same–different test with a 14-month-old.

driven by weaker underlying auditory memory representations (of the unfamiliar nonsense word *bih* or an unfamiliar melody) relative to more-familiar representations (of the familiar word *ball* or a familiar melody like “Twinkle Twinkle Little Star”). According to this *specific-learning hypothesis*, more-familiar items should be both *easier to discriminate* and *easier to associate* than unfamiliar items with similar properties. However, the most supportive evidence for this hypothesis comes from the word-learning literature, and even that evidence is scattered across multiple studies.

The present study

To assess whether the specific-learning hypothesis applies to nonspeech auditory patterns as well as to spoken language, the present study tested preschool-aged children’s abilities to form audiovisual associations with, and to discriminate, either familiar melodies or scrambled (and thus unfamiliar) versions of those melodies. If children were to perform with equal accuracy on familiar and unfamiliar songs, this would fail to support the specific-learning hypothesis. It would also suggest that musical pattern learning operates differently from pattern learning in language, inconsistent with a general-auditory account of pattern learning. On the other hand, better performance on familiar than on unfamiliar songs would lend credence to the specific-learning hypothesis and also support a general-auditory view of pattern acquisition. That is, if the results were to reveal familiar-form effects in nonspeech auditory processing similar to those found previously in spoken word processing, this would suggest that pattern-learning processes are general across multiple types of auditory events, rather than being restricted to spoken language.

Pretesting

The initial pretesting here aimed to determine what melodies children were best at identifying, out of a set of likely candidates. The first pretest tested 19 preschool-aged children by

presenting six childhood melodies (“Mary Had a Little Lamb,” “Twinkle Twinkle Little Star,” “Happy Birthday,” “Deck the Halls,” “Yankee Doodle,” and “London Bridge”) and asking children to name them. After this, the melodies will be referred to by only the first word of the title, for brevity. In the first block of trials, two measures (6–8 notes) of each song were presented once; in the second block, four measures (12–17 notes) of each song were presented once. The melodies were played in a synthesized piano timbre. Accuracy was defined as naming any semantic content in the song, not just the title or lyrics; for example, for “Mary Had a Little Lamb,” the answers “Mary,” “lamb,” and “sheep” were all accepted. Even with this lenient criterion, naming accuracy (Table 1) was fairly low (21% overall) and did not differ between the two-measure and four-measure melodies. The highest naming rates were for “Happy” (36%), “Mary” (33%), and “Twinkle” (39%). However, since children might possess some perceptual knowledge of the songs but have difficulty verbalizing it, a second pretest was conducted.

The second pretest, with 23 children, used a two-alternative forced choice picture-matching task with the same melodies, except that “London Bridge” (0% naming recognition) was replaced with the “Star-Spangled Banner.” The pictures were as described in Table 1. Here, the overall accuracy was 59%. Length (two vs. four measures) did not impact recognition. The pairs with the highest accuracy were “Mary” versus “Twinkle” and “Birthday” versus “Twinkle” (both 67%; note that Table 1 shows per-melody accuracy). Since “Mary” and “Twinkle” have identical rhythmic patterns, this melody pair was selected, allowing for an examination of pitch contour effects in isolation from timing differences. Children’s modest pretest performance on familiar songs will be revisited in the General Discussion.

As was described above, previous studies (Creel 2014c, 2016) showed that children have difficulty distinguishing musical patterns by their pitch contours. Thus, one option for the present study was to test these two familiar melodies and to compare the results qualitatively to those from previous studies. This would parallel work in child language research (Fennell & Werker, 2003; Stager & Werker, 1997) and child music research

Table 1 Pretest stimuli

Melody	Picture	Naming accuracy	Recognition accuracy
Mary Had a Little Lamb	Lamb	.325	.612
Twinkle Twinkle Little Star	Star	.390	.670
Happy Birthday	Cake with candles	.361	.627
Deck the Halls	Christmas tree	.182	.533
Yankee Doodle	US Revolutionary soldier	.000	.536
London Bridge	(not used in Pretest 2)	.000	N/A
Star-Spangled Banner	Flag	N/A	.565

(Corrigall & Trainor, 2010; Trainor & Trehub, 1994), in which findings of sensitivity to familiar versus unfamiliar sound patterns are not compared in the same study. However, comparing the two familiar songs here with the unfamiliar songs from previous studies was less than desirable, because the melodies used in earlier work were shorter and less complex (four or five notes with unidirectional pitch contours) than the melodies used here (seven notes and bidirectional pitch contours). A better unfamiliar control, then, would be melodies with properties more directly matched to the familiar melodies used. Therefore, after selecting the two familiar melodies, a scrambled-note version of each was created (Fig. 1). (Note that the note-scrambling was done at the notation level, so that actual rearranging of the segments of an audio file was not necessary, which could have led to unnatural juxtapositions of reverberation from previous notes.) Simple contours were used for each, so that the familiar and unfamiliar melodies were matched for overall naturalness. Both scrambled melodies ended on the same note as the original melodies, so that the note duration was realistic (final notes of phrases tend to be longer) and so the degree of key resolution was similar to the originals. This use of control unfamiliar stimuli within the same study represents an advance over previous studies in both the word recognition and music perception literatures.

Experiment 1: Association learning

In the first experiment, children were asked to associate two melodies with two different pictures. Since previous studies (Creel, 2016) had suggested that children are unable to associate two melodies that differ only in contour with two different pictured objects, it seemed that the strongest test of melody familiarity effects would come from this difficult task. The interest here was also in whether children would use preexisting associations with familiar melodies, or if familiar-melody benefits extended to both melody-related and novel pictures. Therefore, half of the children were asked to learn that melody-related pictures “went with” familiar or unfamiliar melodies, whereas the other half were asked to learn that *novel* pictures—the cartoon characters used in previous melody–picture association studies (Creel, 2014c, 2016)—“went with” familiar or unfamiliar melodies.

Method

Participants Sixty-four preschool-aged children (ages 3–5) took part. A second, replication sample of 64 children was also obtained. The results are reported together, for clarity and brevity. Five additional children were excluded from the analysis because they did not complete the task.

Stimuli The four melodies were “Mary,” “Twinkle,” “Mary”-scrambled, and “Twinkle”-scrambled (Fig. 1). The melodies were synthesized in piano timbre with the Finale 2009 software (MakeMusic, Inc.) and were edited and scaled to 70 dB SPL in Praat (Boersma & Weenink, 2014). Two different picture sets were used (Fig. 2): either a lamb and a star (familiar-picture condition), or two cartoon characters (unfamiliar-picture condition). The related pictures were chosen to be related to the lyrics of each song, on analogy with Fennell and Werker’s (2003) study using the words *ball* and *doll* and pictures of those objects. The two of the cartoon characters had been used in a variety of studies of word learning (Creel, 2014a, 2014b), talker-voice learning (Creel & Jimenez, 2012), and melody learning (Creel, 2014c, 2016). In all cases in which the learned elements were perceptually distinct, children achieved high rates of learning accuracy (80+%). Thus, if children had difficulty associating the melodies with the cartoon characters, it would not be due to difficulty visually discriminating the cartoon characters.

Procedure Children received reinforced learning trials in blocks of eight trials each (four with one melody–picture combination, four with the other). In both the original and replication studies, 16 children each learned to associate the familiar melodies with related pictures; the scrambled melodies with related pictures; the familiar melodies with novel pictures; and the scrambled melodies with novel pictures. In the related-picture conditions, children were told that they would see a star sing the star song, and a lamb singing the lamb song. In the novel-picture conditions, children were told that they would see one creature sing the star song, and another creature sing the lamb song. On each trial, two pictures appeared on the left and right sides of the screen (side counterbalanced across trials), and then a melody played. The child was asked to select the creature

Fig. 1 Experiments 1 and 2: Melody stimuli

Fig. 2 Experiment 1: Visual stimuli. (Top row) Related pictures. (Bottom row) Novel pictures

who had sung the song. Once they had scored at least 7/8 in a block of learning trials, or completed five learning blocks, they continued to unreinforced test trials.

Results

A logistic regression model in R (R Core Team, 2014) was applied to the data (Fig. 3) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Logistic regression takes into account the fact that accuracy is binomially rather than

normally distributed. The dependent variable was accuracy, with correct coded as 1 and incorrect coded as 0. This model used the predictors melody familiarity (familiar, scrambled), picture familiarity (related pictures, novel cartoons), feedback (reinforced learning trials, nonreinforced test trials), and their interactions. To account for the within-subjects nature of the data, the model included random intercepts for participants and feedback random slopes for participants (the other variables were manipulated between subjects). For feedback trials, the results from just the first block of learning trials—the only block in which all participants took part—were analyzed.

Original sample An effect of feedback emerged ($B = 0.16$, $SE = 0.06$, $z = 2.70$, $p = .007$), suggesting that children became less accurate once reinforcement was no longer available. There was also a Melody Familiarity $\times$ Picture Familiarity interaction ($B = 0.30$, $SE = 0.08$, $z = 3.53$, $p = .0004$), apparently stemming from an advantage for familiar over scrambled melodies that occurred only when the pictures were related. However, this interaction was qualified by a three-way Feedback $\times$ Melody Familiarity $\times$ Picture Familiarity interaction ($B = 0.13$, $SE = 0.06$, $z = 2.35$, $p = .02$). To examine this three-way interaction, the simple two-way Melody Familiarity $\times$ Picture Familiarity interactions were analyzed separately for reinforced and nonreinforced trials. For reinforced (learning) trials, the effect of melody familiarity was significant ($B = 0.24$, $SE = 0.12$, $z = 2.03$, $p = .04$), as was the two-way interaction ($B = 0.47$, $SE = 0.12$, $z = 3.92$, $p < .0001$). Examining melody familiarity at each level of picture familiarity revealed

Fig. 3 Experiment 1 (upper row) and its replication (lower row): Accuracy in the first block of learning trials (left) and in the test trials (right). Dashed lines represent chance performance. Error bars show standard errors. The plots were created in R using ggplot2 (Wickham, 2016)

that only when the picture was related was there a significant benefit ($B = 0.73$, $SE = 0.19$, $z = 3.86$, $p = .0001$) of familiar over scrambled melodies.

Next, the Melody Familiarity $\times$ Picture Familiarity interaction was analyzed for nonreinforced trials. There was no main effect of melody familiarity, but the interaction was significant ($B = 0.20$, $SE = 0.09$, $z = 2.34$, $p = .02$). Looking at each level of picture familiarity separately, the effect of melody familiarity was marginally significant ($B = 0.23$, $SE = 0.12$, $z = 1.94$, $p = .051$) for related pictures, but not significant for novel pictures. No other effects were significant. To restate the interaction pattern: When pictures were related, children were better at associating familiar melodies than at associating scrambled melodies. For novel melodies, no such familiar-melody advantage emerged. The three-way interaction with feedback appears to have come from the fact that the two-way interaction was larger in magnitude for the learning (reinforced) trials.

Replication sample The same analysis was performed on the replication sample. The effect of feedback was marginally significant ($B = 0.12$, $SE = 0.07$, $z = 1.68$, $p = .09$), with slightly lower performance when reinforcement ceased. An effect of picture familiarity ($B = 0.22$, $SE = 0.10$, $z = 2.13$, $p = .03$) was due to higher accuracy for related than for novel pictures. There was also an effect of melody familiarity ($B = 0.30$, $SE = 0.10$, $z = 2.91$, $p = .004$), with higher accuracy for familiar than for novel melodies. No interactions were significant. However, because of the asymmetric effects of melody familiarity in the original sample, melody familiarity effects were examined at each level of picture familiarity. For related pictures, the effect of melody familiarity was significant ($B = 0.47$, $SE = 0.18$, $z = 2.56$, $p = .01$), with stronger performance on familiar melodies. However, for novel pictures, the effect of melody familiarity was not significant.

Discussion

Both the original experiment and the replication suggested that melody familiarity has strong effects on melody encoding, mainly when the association naturally fits with preexisting song knowledge. This closely mirrors effects found in the word-learning literature, in which 14-month-old children respond to differences in the pronunciation of familiar words (*ball* vs. *doll*; Fennell & Werker, 2003) but do not respond to slight differences in the pronunciations of newly learned words (*bih* vs. *dih*; Stager & Werker, 1997). An additional question was also tested: whether the familiar-sound-pattern advantage extends to novel associations. This question has not been addressed in the word-learning literature (such as by labeling two novel objects “ball” and “doll”). The finding here was that the familiar-melody advantage might not extend to novel picture associations. This is especially interesting in

that the instructions pointed out the connection to children even for novel pictures: Children were explicitly told that one creature sang the “star song” and the other sang the “lamb song.” It seems, in principle, that they could simply have remembered one creature as the star singer and the other as the lamb singer, yet they did not seem to do so.

Of course, in both the present study and possible future child word-learning studies, preexisting associations might interfere with learning new ones. That is, knowing what *ball* refers to or having preexisting semantic associations with “Twinkle Twinkle Little Star” might interfere with associating each stimulus with new, unfamiliar visual materials. This appeared to be more the case in the original experiment, in which novel pictures showed a numerically reversed melody-familiarity effect, than in the replication, in which they did not. Given the variability between these two patterns of results, we decline to make a strong statement. It is worth noting that in the child word-learning literature, homophone interference effects (e.g., inability to learn that an unfamiliar object is called a “rope”) appear to surface mainly when the familiar object is present as a response choice (Doherty, 2004). Thus, concerns of existing-association interference may be strongest in cases in which the existing associate is present to interfere with other stimuli. Furthermore, Storkel and Maekawa (2005) reported *better* naming accuracy when children learned that a novel object had a homophonous name like “comb” rather than a nonsense-word name like “bine,” which is consistent with a sound-pattern familiarity effect in word learning.

In any case, the finding here suggests that, unlike in previous studies that had used only unfamiliar melodies (Creel, 2014c, 2016), children can learn melodic contour–picture associations, and do so better for familiar than for unfamiliar melodies. It also raises the question of whether the familiarity advantage is at the level of the perceptual representation, of the learned association (between musical patterns and lyrics or contexts), or both. Therefore, the next experiment was designed to test whether children perform better on familiar melodies in a task that more directly assesses perceptual processing: a same–different task. If children have more robust perceptual representations for familiar than for unfamiliar (scrambled) melodies, then discrimination performance should be better for changes from one familiar melody to another than for changes from one scrambled melody to another.

This experiment allowed for an additional test of potentially heightened salience for familiar melodies. Previous studies of brief-tone sequence discrimination suggested that children are less sensitive to melodic contour than to timbre—that is, to the musical instrument’s sound quality (Creel, 2014c, 2016). Therefore, the next experiment also included timbre discrimination trials, at three different levels of timbre similarity. This

allowed for an assessment of whether differences between *familiar* melodies (vs. the unfamiliar ones tested previously) might be *more* salient than timbre differences.

Experiment 2: Discrimination

Method

Participants Ninety preschool-aged children (ages 3–5 years) who had not taken part in Experiment 1 participated. An additional 51 children took part but were excluded, due to failure to meet the training criterion (29; see Creel, Weng, Fu, Heyman, & Lee, 2018, for similar rates of failure to meet training criterion), noise disruptions at the testing site (five), failure to complete the study (five), unwillingness or inability to follow instructions (two), extreme inattentiveness (one), computer errors (three), or having exposure to a tone language (six), which might change performance on the task (Creel et al., 2018).

Stimuli The melodies were synthesized in multiple timbres using the Finale 2009 software (MakeMusic, Inc.) and were edited and scaled to 70 dB SPL in Praat (Boersma & Weenink, 2014). “Mary,” “Twinkle,” “Mary”-scrambled, and “Twinkle”-scrambled were heard by all participants. The participants also heard different-timbre trials, in which the same melody was played but the timbre changed. For one-third of the participants, the two timbres were very similar (piano and guitar; see Fragoulis et al., 2006), moderately similar (bassoon and alto saxophone), or distinct (vibraphone and muted trumpet). The timbre distinctiveness estimates for the latter two pairs were drawn from Iverson and Krumhansl (1993). For a child in a given timbre condition, the different-contour trials occurred equally often in the two timbres used in that condition. That is, a different-contour trial might contrast “Mary” in

a bassoon timbre with “Twinkle” in a bassoon timbre. No trials contained both contour and timbre differences.

Procedure Children first received training trials on which highly distinct melodies were used (rising vs. falling, high harp vs. low tuba—a 1.5-octave difference in pitch range), including four “same” trials and four “different” trials in each eight-trial training block. If children did not achieve at least seven out of eight correct, the eight-trial training block repeated, up to five total training blocks. Children who never passed the training were excluded from analysis. When the criterion was achieved, the child continued to the test phase. The test phase presented all trials in a random order. The test stimuli included eight melody-change trials (half familiar melodies, half scrambled), eight timbre-change trials (half familiar melodies, half scrambled), and 16 same trials. An additional eight trials presented the training stimuli (half same, half different), to assess continued task adherence. These trials were not analyzed.

Results

Prior to the analysis, accuracy was converted to d -prime, a standard measure of change detection (Macmillan & Creelman, 2005). The extreme values 0 and 1 were converted to $0 + 1/(2N) = .125$ and $1 - 1/(2N) = .875$, respectively, to avoid z scores of $\pm\infty$. The results appear in Fig. 4. An analysis of variance was then conducted on the d -prime values, with the independent variables timbre similarity (close, mid, far; between subjects), melody familiarity (familiar, scrambled; within subjects), and trial type (different melodies, different timbres; within subjects).

An effect of melody familiarity was apparent [$F(1, 87) = 15.85, p = .0001, \eta_p^2 = .15$], suggesting higher d -prime scores for familiar melodies. There was also an effect of timbre similarity [$F(2, 87) = 15.21, p < .0001, \eta_p^2 = .26$], suggesting

Fig. 4 Experiment 2: Discrimination accuracy, with standard errors. The plots were created in R using ggplot2 (Wickham, 2016)

that overall accuracy increased as the timbres became less similar. The timbre similarity main effect was qualified by a Timbre Similarity $\times$ Trial Type interaction [$F(2, 87) = 33.98, p < .0001, \eta_p^2 = .44$], which appeared to result from differences in the relative discriminability of the timbres in the different conditions, such that melody discrimination exceeded timbre discrimination for close timbres, roughly matched it for moderately similar timbres, and undershot timbre discrimination for far timbres. Finally, a Melody Familiarity $\times$ Trial Type interaction [$F(1, 87) = 5.88, p = .02, \eta_p^2 = .06$] suggested that the effect of melody familiarity was larger (.32 vs. .06) and significant [$F(1, 89) = 18.78, p < .0001, \eta_p^2 = .17$] when children were discriminating melodies, but when they were discriminating timbres it was not significant [$F(1, 89) = 0.71, p = .40, \eta_p^2 = .01$].

Discussion

Children appear to be better at distinguishing familiar melodies from each other than at distinguishing unfamiliar melodies from each other. This result held when melodies were played in a variety of timbres. This suggests that children have more robust perceptual representations of familiar than of unfamiliar melodies. Interestingly, a large timbre difference still seems more salient than a difference between melodic contours, even when those melodic contours are familiar. This suggests that melody familiarity does not become more salient than substantial timbre differences, a salience pattern that has previously been reported in this age group (Creel, 2014c, 2016).

General discussion

The original question motivating this study was whether children might perceptually represent music by storing entire melodies, rather than melodic properties, as the specific-learning hypothesis predicts. If so, then familiar melodies should show perceptual processing advantages over matched unfamiliar melodies. This appears to be the case: Children are better at mapping familiar melodies to pictures than they are at mapping unfamiliar melodies to pictures, and they are better at judging perceptual distinctions between familiar melodies than between unfamiliar melodies. These findings are consistent with the specific-learning hypothesis, and more generally with exemplar-style accounts of auditory pattern learning across domains.

The picture-association effect is particularly interesting, in that two previous studies (Creel, 2014c, 2016) showed that children in this age range are unable to learn associations between novel melodic contours and novel pictures. Those findings suggest that pitch-contour differentiation in the context of a learning task is quite difficult. It seems that melody

familiarity is sufficient to ease this task. Of course, one might reverse this observation, to ask why children did not perform better at learning visual associations that were linked to familiar songs' lyrics. One answer may be that not all children were familiar with these songs, and only the ones who were familiar could learn the association. A different answer is that children might be disadvantaged here versus recognizing familiar music in natural situations, because in natural listening situations they have access to additional distinguishing cues, such as timbres and lyrics (see Vongpaisal, Trehub, & Schellenberg, 2009, for evidence that timbres and lyrics contribute to children's music recognition).

In sum, this research contributes to the literature on auditory perceptual development by suggesting that, as in the specific-learning hypothesis, children learn the representations of specific auditory events, rather than simply accruing melodic properties.

Relationship to word learning

The research presented here also speaks to the literature on word learning, by suggesting that the principles governing word learning are in fact broader perceptual-learning and/or associative-learning principles. Recall two contrasting results from child word learning: Whereas pictures with subtly different novel word names (*bih*, *dih*) are not recognized by 14-month-olds (Stager & Werker, 1997), subtly different familiar words (*ball*, *doll*) are recognized easily by the same age group (Fennell & Werker, 2003). In the present study, with slightly older children in an audiovisual association task that is similar to word learning, familiar melodies were more readily associated and recognized than unfamiliar melodies, and in an immediate discrimination task, familiar melodies were more readily differentiated than unfamiliar melodies. The latter finding has not, to my knowledge, been demonstrated in young children's spoken word processing.

An open question is whether the role of familiarity is based not only on familiarity with the auditory form, but also on the existence of semantic associations. Clearly the familiar auditory form is necessary, in that semantic associations did not help in the presence of scrambled melodies. But do children use auditory familiarity alone, or rather some composite of auditory familiarity and semantic associations? In the picture-matching case (Exp. 1), the familiarity effect might depend on the naturalness of the picture mapping itself: Picture associations with "Twinkle Twinkle Little Star" are learned best when the picture is a star, not when it is a novel cartoon character. This suggests that existing semantic associations have an influence on performance. Still, it is interesting that having semantic associations pointed out did not help children learn to associate the melodies with novel pictures. Had children simply associated the word "lamb" with one character and "star" with the other character, this presumably

would have been an easy task for them, yet it was not. See Creel (2014b) on children's high learning accuracy for dissimilar verbal labels for the same cartoon characters.

The similarity judgments are easier to interpret as support for the familiarity of auditory representations, since children could respond without activating meaning associations. Still, one might ask whether children performed better on familiar melodies in the discrimination task because they recoded the real melodies into meaning-based representations, which they could not do for the unfamiliar melodies. This is certainly possible. However, it should be kept in mind that children were not very good at naming the songs in the pilot data, even when prompted to do so (accuracy of about 35% for the two selected songs), suggesting that lyrics recall, and therefore semantic associations, may not be automatic for children. Given that the children in Experiment 2 were not told that they would be hearing familiar melodies, it is possible that many or most of them did not register that some songs were familiar.

To the extent that the semantic account of the present results is valid, it may suggest a different parallel between word-form learning and auditory pattern learning. Specifically, the semantic account may indicate that external (semantic or lyrical) associations facilitate the learning of auditory form, an account that has been proposed for word learning (e.g., Yeung & Werker, 2009). That is, perhaps association with distinct semantic networks sharpens or pattern-separates the auditory representations of melodic properties.

Regardless of the exact nature of familiar-melody facilitation, though, it is clear that long-term familiarity of some sort—that is, learning melodies and their associations over multiple days, weeks, or months—is necessary to producing these effects. Creel (2014c, 2016) found little learning with multiple exposures in brief lab association experiments. Furthermore, Creel's (2016) Experiment 1 presented children with extra exposure to the novel melodies prior to the association task, but even this extra exposure did not allow children to learn associations. Thus, extensive exposure, but not brief exposure, to specific auditory patterns may be necessary to produce facilitated performance.

Limitations

A limitation to this study is the small number of melodies used. It remains possible that subtle differences between the familiar and unfamiliar melodies may have driven differences in the effects. As was discussed with the preliminary studies, the use of a large number of melodies was constrained by the number of melodies children can reliably recognize, and they had difficulty recognizing these ostensibly familiar melodies. Although many child language experiments depend on a relatively small number of words, such as the foundational Stager and Werker (1997) study, it would be reassuring to see the present effect extended to other data points. One

fruitful approach would be to test children in a single school setting in which the children's musical exposure repertoire was well-known, or, even better, in two different schools where children learned two different sets of melodies, providing a fully crossed design. Such work should include a wider range of children's music, especially music without lyrics and without differing extramusical associations, in order to dissociate roles of semantic association and auditory familiarity.

A second limitation is that the present research does not make clear what properties children can use to organize the representations of familiar melodies. In the present study, melodies with identical timing properties were carefully selected, so that the melodic patterns were the focus of inquiry. Of course, melodic patterns include not just contour, but exact pitch distances, scale degrees included, pitch range, and tonality (major/minor, which did not vary here). The results here show that well-known melodies themselves, not just schematic melodic features, may be an organizing principle. Future work should examine additional factors in children's musical representations.

Author note Data and R code for analysis are available at the following DOI: 10.17605/OSF.IO/WH3FY.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 5.4.01) [Computer program]. Retrieved November 9, 2014, from www.praat.org
- Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, 28, 195–200.
- Creel, S. C. (2014a). Impossible to ignore: Word-form inconsistency slows preschool children's word-learning. *Language Learning and Development*, 10, 68–95. <https://doi.org/10.1080/15475441.2013.803871>
- Creel, S. C. (2014b). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*, 73, 81–98. <https://doi.org/10.1016/j.jml.2014.03.001>
- Creel, S. C. (2014c). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1146–1160. <https://doi.org/10.1037/a0036057>
- Creel, S. C. (2016). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science*, 14, 373–303. <https://doi.org/10.1111/cogs.12237>
- Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, 113, 487–509.
- Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in spoken-language processing? *Trends in Cognitive Sciences*, 19, 713–716. <https://doi.org/10.1016/j.tics.2015.09.006>

- Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental Science*, 21, e12503. doi:<https://doi.org/10.1111/desc.12503>
- Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, 85, 2154–2164. <https://doi.org/10.1121/1.397864>
- Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, 31, 203–214. <https://doi.org/10.1017/S030500090300583X>
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354. <https://doi.org/10.1037/0033-295X.85.4.341>
- Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46, 245–264. <https://doi.org/10.1177/00238309030460020901>
- Fragoulis, D., Papaodysseus, C., Exarhos, M., Roussopoulos, G., Panagopoulos, T., & Kamarotos, D. (2006). Automated classification of piano–guitar notes. *IEEE Transactions on Audio, Speech, and Language Processing*, 14, 1040–1050. <https://doi.org/10.1109/TSA.2005.857571>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>
- Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122. <https://doi.org/10.3758/BF03194458>
- Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, 94, 2595–2603. <https://doi.org/10.1121/1.407371>
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195–1197.
- Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153–169.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, 1, 272–276. <https://doi.org/10.1111/j.1467-9280.1990.tb00213.x>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.
- McFadden, D., & Callaway, N. L. (1999). Better discrimination of small changes in commonly encountered than in less commonly encountered auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 543–560. <https://doi.org/10.1037/0096-1523.25.2.543>
- Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1377–1399. <https://doi.org/10.1037/xlm0000247>
- R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381–382.
- Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, 32, 827–853. <https://doi.org/10.1017/S0305000905007099>
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484. <https://doi.org/10.1111/1467-9280.00485>
- Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, 56, 125–132. <https://doi.org/10.3758/BF03213891>
- Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, 55, 821–830.
- Vongpaisal, T., Trehub, S. E., & Schellenberg, E. G. (2009). Identification of TV tunes by children with cochlear implants. *Music Perception*, 27, 17–24.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3, 1–30. <https://doi.org/10.1207/15250000252828226>
- Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113, 234–243. <https://doi.org/10.1016/j.cognition.2009.08.010>