UC San Diego
UC San Diego Previously Published Works

Title

Ups and Downs in Auditory Development: Preschoolers' Sensitivity to Pitch Contour and Timbre

Permalink

https://escholarship.org/uc/item/8xf2m2ng

Journal

Cognitive Science, 40(2)

ISSN

0364-0213

Author

Creel, Sarah C

Publication Date

2016-03-01

DOI

10.1111/cogs.12237

Peer reviewed

Ups and Downs in Auditory Development: Preschoolers' Sensitivity to Pitch Contour and Timbre

Sarah C. Creel

Department of Cognitive Science, University of California San Diego

Abstract

Much research has explored developing sound representations in language, but less work addresses developing representations of other sound patterns. This study examined preschool children's musical representations using two different tasks: discrimination and sound–picture association. Melodic contour—a musically relevant property—and instrumental timbre, which is (arguably) less musically relevant, were tested. In Experiment 1, children failed to associate cartoon characters to melodies with maximally different pitch contours, with no advantage for melody preexposure. Experiment 2 also used different-contour melodies and found good discrimination, whereas association was at chance. Experiment 3 replicated Experiment 2, but with a large timbre change instead of a contour change. Here, discrimination and association were both excellent. Preschool-aged children may have stronger or more durable representations of timbre than contour, particularly in more difficult tasks. Reasons for weaker association of contour than timbre information are discussed, along with implications for auditory development.

Keywords: Auditory development; Music perception; Auditory memory; Memory development; Perceptual learning; Word learning; Timbre

1. Introduction

How do children represent the sound patterns they hear? A great deal of research has explored representations of sound in language, but less is known about sound patterns in other domains. A full understanding of how sound pattern representations form and change is crucial to questions about the role of perceptual learning and maturation in memory formation.

In language, children must come to map word forms—the sound patterns of words—to event representations and to referents in their environments. Furthermore, children must sort out the subset of variability in words' sound patterns that corresponds to meaning changes. For instance, children must learn that a male talker saying *dog* is equivalent to a female talker saying *dog*, which is different than the female talker saying *doll*. Thus, voice gender does not change meaning, but changing the sound /g/ to the sound /l/ does.

A somewhat parallel phenomenon occurs in melody recognition, albeit for a different set of auditory properties. Specifically, adults regard a melody's identity as being determined by the specific pitch intervals between successive notes (as well as timing patterns). That is, they know that Happy Birthday played by a kazoo is the same melody as Happy Birthday played by a tuba (an instrument with a different timbre), which is in turn different from Jingle Bells played by a tuba. Thus, an instrument change does not change what melody it is, but a change in pitch order and pitch content does.

One might ask whether word learning in language and in music have anything in common, in that word learning entails forming associations of words with referents, while it is less clear what music refers to. Scholars have long argued about whether music does, or even can, refer to anything in the outside environment (see Locke, 1986; for a review). Recent work suggests some culturally shared extramusical associations (e.g., Eitan & Timmers, 2010; Fritz et al., 2009), and further that some affective states evoked by music (the major/minor key distinction) may be learned by association (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). These studies suggest that listeners form—or in some cases may already possess—associations between music and non-musical concepts.

Even more important for this study, part of children's social enculturation is to associate different musical or music-like events with different aspects of the environment. That is, while children do not learn that musical words *refer to* objects or events in their environments, they do need to learn to individuate melodies and (as with other non-speech environmental sounds) associate them with different environmental concomitants. These include culturally relevant events (birthday parties vs. holiday gatherings), door bells, cell phone ring tones (Roye, Jacobsen, & Schröger, 2007), television show theme songs (Vongpaisal, Trehub, & Schellenberg, 2009), radio station identifications, car alarms, or affective associations in cinematic music. In language, children must also learn mappings between pitch patterns both with affective states (happy, sad, angry; see, e.g., Morton & Trehub, 2001) and with sentence type (such as "They ate ice cream." vs. "They ate ice cream?"). Thus, it is reasonable to think that children are accustomed to experiencing music or pitch patterns being associated with other events. Furthermore, learning and using such associations will necessarily be constrained by children's abilities to individuate one melody from another. It is that individuation that the current paper focuses on.

For language, research suggests that children sort out relevant from irrelevant variability fairly early in life. However, not as much is known about children's representations of relevant versus irrelevant perceptual aspects of music. This study aimed to assess young children's ability to process musical materials. How well do their representations retain pitch contour information as opposed to "surface" features such as instrumental timbre—do they, as in language, sort out relevant from irrelevant variability early on, or

are they still learning what the musically relevant information is at an age when language input is already correctly weighted?

1.1. Previous research on sound pattern recognition

Studies of language processing have been devoted to understanding how and when children determine that speech sound content, and not talker characteristics, indicate a word's identity. This research indicates that children do learn that talker variation is largely irrelevant to word identity. However, this learning process is not instantaneous. Early in development, 7.5-month-olds do not recognize familiarized words over a change in talker gender (a large acoustic change, but not a phonemic[1] change; Houston & Jusczyk, 2000). By 10.5 months, infants do recognize words over a talker gender change (Houston & Jusczyk, 2000).

Later, as children begin learning word-meaning mappings, they map similar-sounding words (e.g., *bih* and *dih*) spoken by a single talker to different visual objects by 17 months (Werker, Fennell, Corcoran, & Stager, 2002). Interestingly, Rost and McMurray (2009, 2010) found that children succeed earlier in this similar-word-learning task, at 14 months, if each word is spoken by a variety of talkers—that is, children hear variability on a non-criterial "surface" dimension. Rost and McMurray hypothesized that, at 14 months, children need assistance (in the form of variability) in ruling out talker variation as a relevant property of word representations. Older children appear not to need such scaffolding: Research (e.g., Creel, 2014) suggests that by around 4 years, if not earlier, children can learn two similar words (e.g., *geeb* and *geege*), each spoken by a different talker, and readily transfer recognition when the word is spoken by the other talker. That is, children do not strongly associate talker-specific speech cues to referents—at least, not to the point that recognition accuracy is substantially impaired—even though there is a consistent within-experiment pairing between a talker's voice and a pictured referent. Nonetheless, talker information may be residually present even in adults' word representations or episodic traces (Creel, Aslin, & Tanenhaus, 2008; Goldinger, 1996, 1998; see Creel, 2014, for similar effects in preschoolers), suggesting that listeners still register talker variation when forming word representations. This implies perhaps that the downweighting of talker information is not absolute.

Outside of speech processing, the contents of children's auditory memory representations are less certain. Pioneering work by Trehub and collaborators suggests that infants as young as 5 months (Chang & Trehub, 1977; Trehub, Bull, & Thorpe, 1984) distinguish brief melodies differing in contour. At 9–11 months, infants can detect a small, contour-preserving change to a single tone in a brief melody, though they do not show facilitated change detection for *culture-specific* musical contexts until age 4–6 years (Trehub, Cohen, Thorpe, & Morrongiello, 1986). Adults are more sensitive to the musical scales and harmonic patterns of their culture than 5-year-olds are (Trainor & Trehub, 1994). Nonetheless, adults remain highly sensitive to contour, often confusing non-identical same-contour melodies for each other (Dowling, 1978; Massaro, Kallman, & Kelly, 1980), and experiencing cross-modal mappings between contour and physical movement

(Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). Thus, contour sensitivity appears to emerge very early in development and exerts a continuing influence on the organization of musical memory, with infants' initial sensitivity to pitch progressively tuned toward music of their culture (see Lynch, Eilers, Oller, & Urbano, 1990; Trehub et al., 1986), analogous to the "language-general-to-language-specific" progression described in the development of speech perception (e.g., Werker & Tees, 1984).

Though relative pitch (pitch contours and pitch intervals) is perhaps the best-studied musical attribute, infants and children are also demonstrably sensitive to numerous other musical properties, including timing (Hannon & Johnson, 2005; Hannon & Trehub, 2005a,b) and pitch height (infants: Saffran & Griepentrog, 2001; though see Plantinga & Trainor, 2005; for counterevidence; young children: Trehub, Schellenberg, & Nakata, 2008; adults: Levitin, 1994; Schellenberg & Trehub, 2003). They are also sensitive to *timbre*—essentially, differences in the instrument or sound source producing a musical note. Infants detect changes in timbre to tones (Trehub, Endman, & Thorpe, 1990) or familiarized melodies (Trainor, Wu, & Tsang, 2004). Vongpaisal et al. (2009) found that typically hearing 4–6-year-olds recognized familiar cartoon music more readily when the music was presented in its original timbre than when it was presented in a neutral timbre, though they exceeded chance even when timbre cues were neutralized. Adults' melody recognition is also facilitated by timbre (Halpern & Müllensiefen, 2007; Radvansky, Fleming, & Simmons, 1995; Radvansky & Potter, 2000; Schellenberg, Iverson, & McKinnon, 1999), much like recognition of repeated words is facilitated by talker information (Palmeri, Goldinger, & Pisoni, 1993). In summary, children's recognition of music may be supported by a variety of cues, all of which are retained to some degree into adult listening.

1.2. Representations in language and music

One might conclude from the above review that young children and even infants have relatively adult-like musical representations, aside from culture-specific exposure to a musical system (Lynch et al., 1990; Trehub et al., 1986). However, this assumes that high *sensitivity* in an immediate-memory paradigm reflects strong, long-lasting memory representations. It also assumes that implicit responses by infants are homologous to explicit responses by adults. Some studies, both in language development and pitch processing, suggest otherwise.

In language development, at least two lines of research indicate that good immediate processing of sounds does not necessarily entail the ability to form associations. First, research by Stager and Werker (1997) showed that, while infants can detect changes between the novel words "bih" and "dih" (distinguished by a change in a single speech sound) at 8 months, much-older 14-month-old infants are unable to associate those two sound patterns to different pictures. This does not appear to result from general difficulty encoding associations: 14-month-olds readily associate *dissimilar*-sounding words ("lif" vs. "neem") to two pictures (Stager & Werker, 1997). It is not until 17 months that infants can form word-picture mappings with similar-sounding words (Werker et al.,

2002; though recall that Rost & McMurray found good performance at 14 months when talker variability was present; see also relevant research by Fennell & Waxman, 2010; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009, on other conditions under which 14-month-olds can learn audiovisual associations between similar words).

A second line of research in language development by Horst and Samuelson (2008) investigated the phenomenon of "fast mapping" (Carey & Bartlett, 1978), whereby young children form initial associations between a word form and a meaning with a very limited number of exposures. Horst and Samuelson found that, although 24-month-olds form such associations reliably, these associations did not persist over a 5-min delay. Both these lines of research from language development suggest that children's performance in immediate-memory tasks may not represent stable association learning.

A few studies on pitch processing suggest that pitch memory may not be as durable for children as for adults. These studies have found declines in children's memory over the course of a few seconds (Gomes et al., 1999; Keller & Cowan, 1994; see also Trehub et al., 1984; who cite pilot data suggesting that a lengthy interstimulus interval [ISI] precipitously lowered performance in an infant conditioned head-turn paradigm). In a behavioral study, Keller and Cowan (1994) showed that children aged 6–7 years, compared to adults, showed a faster accuracy decline in a pitch change detection task with a variable ISI. This apparently faster memory degradation was evident even though the authors controlled for level of pitch discriminability at the 2-s baseline ISI. A related study by Gomes et al. (1999) replicated this basic result by testing the duration of children's electroencephalogram response to pitch mismatch, the mismatch negativity (MMN). The MMN can be detected even under inattentive conditions. Under inattentive listening conditions (participants heard tones in the background while watching a movie or reading a book), Gomes et al. found that 6–10-year-old children showed an MMN at a 1-s ISI but not an 8-s ISI, whereas 11–12-year-olds and adult listeners showed MMNs at both ISIs. These studies together suggest that children as old as age 10 may experience faster memory degradation for auditory information than adults do.

To summarize, studies of word learning and of pitch discrimination suggest that sensitivity to auditory properties, or auditory pattern-visual pattern mapping in the short term, does not imply robust storage of those properties or associations. Thus, while children are *sensitive* to a variety of musical properties, including pitch contour, it is not certain how rapidly children form robust representations of musical patterns such as melodies, or what these representations contain. Furthermore, it is open to question whether listeners are aided in extracting crucial melody-identifying properties like pitch contour by experiencing variability on irrelevant dimensions (similar to Rost and McMurray's [2009, 2010] finding that irrelevant talker variation facilitates word-meaning mapping in 14-month-olds; but see Plantinga & Trainor, 2005).

The question of memory formation is not one that has been addressed frequently in developmental music processing. Nonetheless, a handful of studies have documented very early abilities to recognize familiar music in infants (Plantinga & Trainor, 2005; Saffran, Loman, & Robertson, 2000; Trainor et al., 2004) and young children (Corrigall & Trainor, 2010; Trehub et al., 2008; Vongpaisal et al., 2009). For instance, Plantinga and Trainor

asked parents of 6-month-olds to play a CD containing six repetitions of an instrumental rendition of a folk song for seven successive days. Children tested in the laboratory on Day 8 showed a novelty preference, listening longer to a novel song than the familiarized song. However, they did not show a novelty preference for a pitch-shifted version over an original-pitch version of the familiarized song, suggesting that song recognition generalized over a change in absolute pitch. In a similar paradigm, Trainor et al. (2004) found that infants detected changes in timbre and tempo (rate of presentation) of highly familiarized music. Vongpaisal et al. (2009) studied recognition of familiar cartoon show songs in typically hearing 4–6-year olds (who served as a control group for older children with cochlear implants). Each child viewed pictures of four cartoon characters that were highly familiar to that particular child. On each trial, the theme song from one of the cartoon shows was played, and children were asked to select the pictured character(s) from that show. Vongpaisal et al. found that the normal-hearing children recognized songs well above chance levels, even when songs' melodies were presented in a neutral timbre (flute, instead of the sung lyrics). However, they were more accurate when they heard the original instrumental timbres. Thus, these studies suggest that children can form detailed long-term musical memories with extensive, distributed exposure. However, these studies do not assess how *readily* different properties of music are associated with other information (e.g., a cell phone, a character in a cartoon show, a favorite television theme) in memory.

1.3. This study

The main purpose of this study was to assess what types of information children apprehend about the music they hear. To assess learning of associations with musical materials, a sound–picture association paradigm was used. This paradigm is similar to paradigms employed to test children's abilities to learn words (e.g., Creel, 2012; see related, but not identical paradigm used by Stager & Werker, 1997; Werker et al., 2002) and voices (Creel & Jimenez, 2012) and is quite similar to Vongpaisal et al.'s (2009) task where children heard cartoon show theme songs and were asked to select a picture of character (s) from that show. Children were introduced to two cartoon characters who, they were told, each had a "favorite song." Children saw each creature and heard its favorite song multiple times before their favorite song knowledge was tested.

The first two experiments tested memory for pitch contour. Two melodies were constructed with very simple, schematic pitch contours: One was a succession of rising pitches, and the other was a succession of falling pitches. On analogy to language, this was akin to presenting two different words (e.g., "lif" and "neem") in the same voice. The third experiment tested memory for timbre. Two different instruments played the same series of rising pitches, and each series was associated with a different cartoon creature. On analogy to language, this was like hearing a female voice label a picture "lif," whereas a male voice labeled another picture "lif." Both Experiments 2 and 3 explored the relationship between children's immediate sensitivity to musical properties (same-different discrimination) and the stability/durability of memory associations (association learning).

An additional purpose of the study was to explore the facilitative effects of auditory experience in association formation, analogous to Rost and McMurray (2009, 2010). Does prior experience with sound patterns facilitate encoding of those patterns, as in the lengthy-familiarization studies cited above (Plantinga & Trainor, 2005; Saffran et al., 2000; Trainor et al., 2004; Vongpaisal et al., 2009)? Furthermore, *variable* exposure may facilitate learning. While variability on non-criterial dimensions has been found to facilitate word learning in toddlers (Rost & McMurray, 2009, 2010), its role in music learning is less clear. Experiment 1 specifically explored the role of presence and type of preexposure on memory formation. Experiments 2 and 3 contained preexposure that directed attention to melodic differences by asking children to make same-different judgments.

The experiments taken together assess how readily children encode particular properties of musical patterns. If studies of word learning and talker variability are taken as a guide, the preschool-aged children tested here should encode contour—as a criterial feature to melodic identity—with perhaps some residual encoding of timbre. On the other hand, it may not be safe to take word learning studies as a guide here, given presumably large differences in children's amounts of attentive exposure to, and their very different motivational and attentional dispositions toward, language versus music. On that view, it is less clear what to expect. Will children encode both melodic contour and timbre readily? Studies tapping immediate memory in infants (e.g., Chang & Trehub, 1977; Trainor et al., 2004; Trehub et al., 1984, 1990) suggest that both attributes may be robustly represented. However, studies of children's memory duration (Horst & Samuelson, 2008; Keller & Cowan, 1994; Stager & Werker, 1997) hint that immediate memory may not guarantee robust association learning.

2. Experiment 1

This experiment looked at whether children can map different-contour melodies to pictures, and whether such association learning is facilitated by preexposure to the melodies. Previous research suggests that preschool-aged children can learn contour and scale-degree (interval) information with massive preexposure (Corrigall & Trainor, 2010; see also Plantinga & Trainor, 2005; Trainor et al., 2004). In language research, at least some types of within-category variability are thought to be especially facilitative (Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008; see also Rost & McMurray, 2009, 2010). Therefore, equal numbers of children heard a brief preexposure sequence containing both melodies at (1) one pitch height, (2) multiple-pitch heights, or (3) no preexposure.

If children can associate different-contour melodies with different pictures, then learning performance should exceed chance overall. If preexposure to melodies facilitates encoding, then preexposed children should map more accurately than unexposed children. If variable exposure in particular aids encoding by highlighting the critical difference between the two melodies, then high variability-exposure children should outperform low variability-exposure children.

2.1. Method

2.1.1. Participants

Twenty-four 3–5-year-old children (11 female; $M = 4.3$, $SD = 0.7$) from local San Diego preschools and day cares took part. Parents were asked to report which language(s) their child hears, speaks, and/or understands. In the current experiment, eight children were reported as hearing another language in addition to English, but none were tone languages. (As discussed later, tone languages have been suggested to sharpen pitch perception abilities, e.g., Pfordresher & Brown, 2009.) Parents were not asked to provide ethnicity data, but for reference, the demographic makeup of San Diego county includes 47% White, 33% Hispanic, 12% Asian, and 6% Black or African American individuals (http://quickfacts.census.gov/qfd/states/06/06073.html). No teachers or parents reported that children had hearing difficulties, and experimenters did not notice comprehension difficulties among the final sample. Three more children were tested but not analyzed due to earphone problems (1), failure to follow instructions (1), or computer problems (1). Recruiting and testing procedures were approved by the UCSD human research protections program.

2.1.2. Stimuli

The two melodies (Fig. 1) and the familiarization sequences were notated in Finale 2009 software (MakeMusic, Inc.: Boulder, CO, US) and were exported to sound files in a MIDI flute timbre at a quarter note duration of 556 ms. Pilot data indicated that a single contour change (e.g., a difference in one note at the end of the melody) was extremely difficult for children to learn. Melodies were therefore calculated to have few notes (4–5) and to have simple contours that were maximally different: One rose in pitch throughout, and one fell in pitch throughout. To provide an additional distinguishing cue between melodies, there was a small durational difference: The rising melody contained five notes whose durations were 139-139-139-139-278 ms, respectively, whereas the falling melody contained four notes whose durations were 278-139-139-278, respectively. As will be evident later in the paper, this durational difference did not facilitate performance.

The *single-pitch* familiarization sequence included 36 repetitions of each of the two melodies, at the pitch level to be used in the sound–picture association task (F5–C6; C4 = middle C). The *multiple-pitch* version also included 36 repetitions of each melody, but here, each melody was heard 18 times at the to-be-learned pitch level (F5–C6), nine

Fig. 1. Rising (left) and falling (right) melodies used in Experiment 1.

times at a perfect fifth above the to-be-learned level (C6–G6), and nine times at a perfect fourth below (C5–G5). Visual stimuli were two cartoon creatures which have been used previously in sound–picture association studies (Creel, 2012, 2014; Creel & Jimenez, 2012).

2.1.3. Procedure

There were two phases of the experiment (Fig. 2): preexposure and sound–picture association (training and test). In preexposure, children were told to watch a butterfly picture which moved up and down on the screen and to listen to the music (93 s). The picture moved slowly (each up/down cycle took roughly 5 s) and was not timed to be synchronous or asynchronous with the music. There is a possibility that children might form butterfly melody associations and that these might interfere with learning melody-creature associations. However, this seemed relatively unlikely on at least three counts: Children at age 3 years show evidence of having learned the same word-form for two different referents (Backscheider & Gelman, 1995), suggesting they might similarly learn one melody for both the butterfly and one of the creatures; the butterfly was not a response option on test trials; the butterfly would be equally associated with both melodies. Thus, on balance, this potential danger seemed smaller than a danger of loss of attention due to lack of visual stimuli. In the no-exposure condition, children began with the association task.

The association task contained training and testing phases. Just before training, children were told that they would see two creatures, each of which had a favorite song. On each training trial, children saw one creature move onto the computer screen and pause in the center. Next, the favorite song (one of the two melodies) played. Then the creature moved offscreen. After eight training trials (four per creature), there was a brief animation (moving animal pictures paired with cheering/clapping sounds) to maintain child interest, which lasted a total of 12.6 s. Eight more training trials (four per creature) followed.

Between training and test, instructions were read to the child. A timed reading of instructions took 8 s, but the time interval between training and test might be longer when running a child participant to allow the experimenter to verify that the child understood and was ready to begin. Then, on each of 16 test trials, both creatures appeared side by side (left picture centered at 25% of screen width, right picture at 75% of screen width, both 20% from top of screen). After 500 ms, one of the melodies played. The child was asked to point to the creature who was singing its favorite song, and the experimenter recorded the child's response with a mouseclick. Left-right location was counterbalanced across trials.
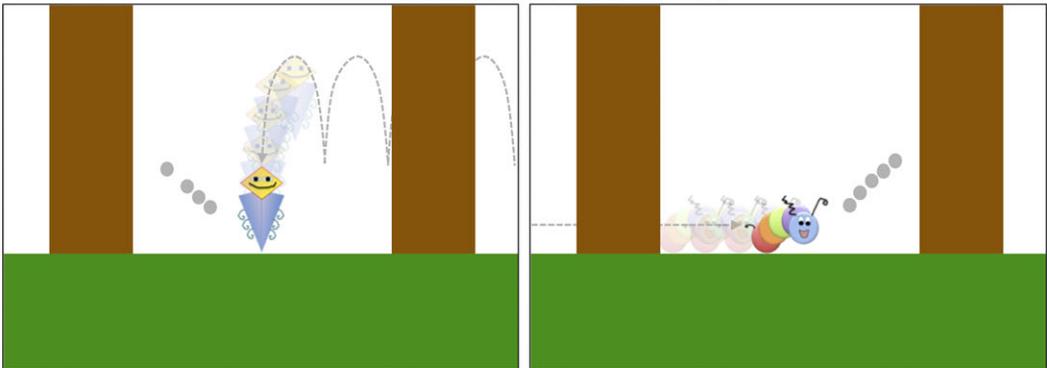
2.2. Results

Counter to the hypothesis that preexposure would facilitate association performance, accuracy across conditions was uniformly poor (Fig. 3, left). As proportion data are standardly transformed to correct for non-normality, data throughout were empirical-logit
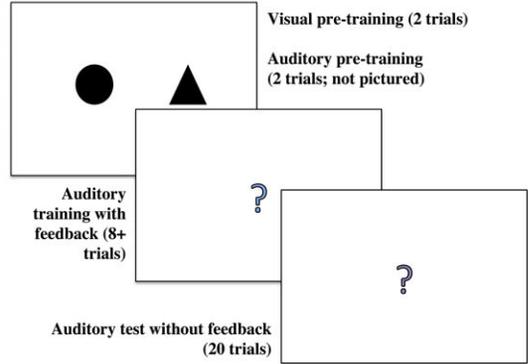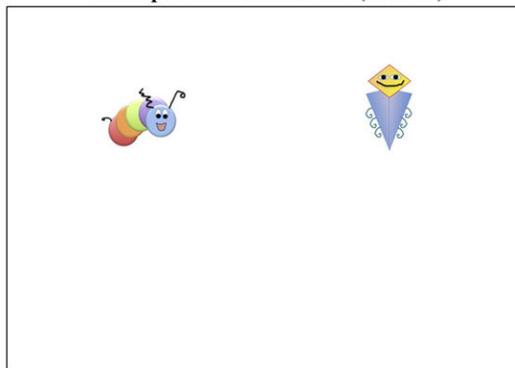
Fig. 2. Schematic of phases in experiments, with example displays. Dashed lines indicate movement paths. Filled circles (training) denote melody played on an example trial. The original butterfly photograph could not be sourced; a similar photograph appears here (Bernard Dupont, flickr.com, Creative Commons licensed for noncommercial reuse).
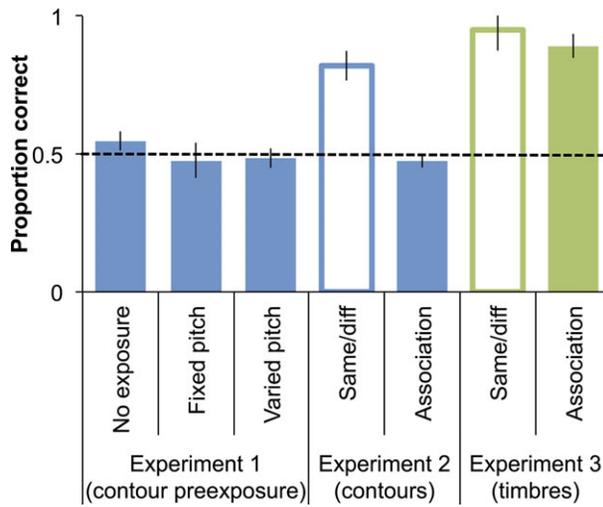
Fig. 3. Accuracy across experiments, with standard errors. Same/diff = response accuracy in same/different task. Association = accuracy in audiovisual association (identifying the cartoon character corresponding to the melody heard). Dashed line = chance responding.

transformed prior to analysis—though, as described in Appendix A, all analyses yielded identical significance patterns for raw data, transformed data, and d-prime scores. Note that, throughout, data in figures and in the text are described in terms of raw accuracy for ease of interpretation. Because the age range tested throughout (most children fell in the age range 4.0–5.5 years) was not as broad as those tested in studies that have found age differences in music processing (e.g., Dalla Bella et al., 2001; ages 3–8 years; Stalinski, Schellenberg, & Trehub, 2008 ages 5–11 years), effects of age on performance were not expected. Age effects are reported for thoroughness.

An analysis of variance (ANOVA) on transformed accuracy with preexposure (none, single-pitch, multiple-pitch) as the between-participants variable did not approach significance ($F(2,21) = 0.58$, $p = .57$, $\eta^2 = .05$). Overall performance did not differ from chance ($t(23) = 0.01$, $p = .99$, Cohen's $d = 0.00$), and none of the individual means differed from chance (no preexposure: $54.7 \pm 9.8\%$; single-pitch: $47.7 \pm 17.9\%$; multiple-pitch: $48.4 \pm 9.8\%$). Furthermore, examination of individual children's performance suggested that none of the children approached perfect accuracy; the maximum score was 12/16 (75%), and the minimum was 2/16 (12.5%). Both scores occurred in the single-pitch preexposure condition. In this and following experiments, there were no indications of significant changes in performance from the first half of trials to the second half of trials. The age-accuracy correlation did not approach significance.

2.3. Discussion

The results here suggest that, even with numerous preliminary exposures to the auditory stimuli, children have difficulty associating different pitch contours to pictures. This

may occur despite previous indications that children form robust music representations (Corrigall & Trainor, 2010) because children need more extensive exposure—distributed over days or months—to form representations of melodic contour- or scale-degree information that are robust enough to allow association learning.

Children's difficulty in melodic association learning implies that memory for newly learned melodies, or the association task, or both, mute the differences between two melodies. Another possibility is that contour differences are not very salient to children in this age group, predicting that children would also show difficulty *discriminating* different contours. Failure on a same-different discrimination task might seem unlikely in light of previous research suggesting good discrimination in infants (e.g., Trehub et al., 1984). However, infants' presumably implicit contour discrimination may not be reflected in overt behavior. That is, children may have implicit awareness but fail to show it in an explicit task (see Corrigall & Trainor, 2014, for a recent illustration of implicit sensitivity without explicit awareness of Western tonality in 4-year-olds).

Therefore, the next experiment replicated the sound–picture association task in a new sample of children. To quantify sensitivity to contour differences, the learning task was paired with a same-different discrimination task. If children are simply insensitive to contour, then accuracy should be low on both discrimination and association. However, if children have good sensitivity to but weak memory for contour, they should perform well on discrimination but poorly on association. Finally, if there is variance in association ability based on individual sensitivity to contour, then discrimination accuracy should be correlated with association accuracy.

3. Experiment 2

3.1. Method

3.1.1. Participants

Twenty-four 3–5-year-old children (16 female; $M = 4.8$, $SD = 0.5$) recruited from schools similar to those in Experiment 1 took part. Most children (23/24) heard languages in addition to English in the home or in intensive classes, and of these, 16 had exposure to a tone language (13 Vietnamese, 2 Mandarin, 1 Cantonese).[2] Six more were tested but not included due to lack of understanding of example same-different trials (4) and computer error (2).

3.1.2. Stimuli

Sounds used during the discrimination task consisted of four highly discriminable training stimuli plus the critical melodies. Training melodies were rising and falling melodies played on either a harp or tuba. Different *training* trials paired together either a rising harp melody starting on F#4 versus a falling tuba melody starting on G3; or, a falling harp melody on C#5 versus a rising tuba melody on C3. That is, the different training trials differed in timbre (harp vs. tuba), pitch (1.5 octaves), and contour (rising vs. falling).

Different *critical* trials were presented in the same timbre (bassoon), and differed only in contour (one rising, one falling). Same trials repeated exactly the same melody (both timbre and contour matched). The only changes in critical melodies relative to Experiment 1 were in tempo (quarter note duration 800 ms, vs. 556 ms in Experiment 1), timbre (bassoon, vs. flute in Experiment 1), and pitch range (F#4-C#5, vs. F5-C6 in Experiment 1).

3.1.3. Procedure

For the discrimination task, the experimenter read the following instructions to each child: "We're going to play the SAME GAME! You get to say whether TWO THINGS are THE SAME, or if they are DIFFERENT. Let's try one." The discrimination task began with two visual example trials, one "different" (a circle and a triangle) and one "same" (two circles). Two example auditory trials followed (one same, one different). Next, children heard a block of eight training trials (four same, four different) using the highly discriminable training stimuli. They received verbal feedback on correctness ("Good job!" or "No, those were different/same"). Children had to answer at least seven of eight training trials correctly in one block to continue to the test. Otherwise, the training block was repeated until they succeeded, or until 3 blocks (24 trials) had elapsed. They then heard 20 discrimination test trials: four "different" trials identical to those in training; four "same" trials identical to those in training; four "same" trials with the critical melodies (two trials with a rising melody presented twice; two trials with a falling melody presented twice); and eight "different" trials with the critical melodies (four with a rising melody and then a falling melody; four with falling, then rising). All auditory trials presented a question mark (color randomly varying) in the center of the screen to provide some visual interest (see Fig. 2).

Note that there were intentionally more "different" trials overall than "same" trials. The reasoning was that, if children had moderate difficulty discriminating the "different" critical melody pairs, they might (subjectively) perceive a greater proportion of same trials than different trials and develop a strong "same" bias. Thus, injecting a few more "different" trials aimed to offset this potential perceived imbalance. Throughout, children were prompted to give a verbal response of "same" or "different."

Discrimination test trials were presented in a fixed quasi-random order. Fixing the order eliminated more than three same or three different responses in a row, and prevented any child from having an advantage due to a guessing strategy working better for one order than another. Children then completed the association task, with procedure identical to Experiment 1.

3.2. Results

3.2.1. Discrimination

Most children (20) passed the reinforced training trials in one block; two children took two blocks; two children completed three training blocks without reaching criterion. All children were included in analyses for comparability with Experiment 1, where there was no way to exclude poor performers. Discrimination accuracy was calculated based on
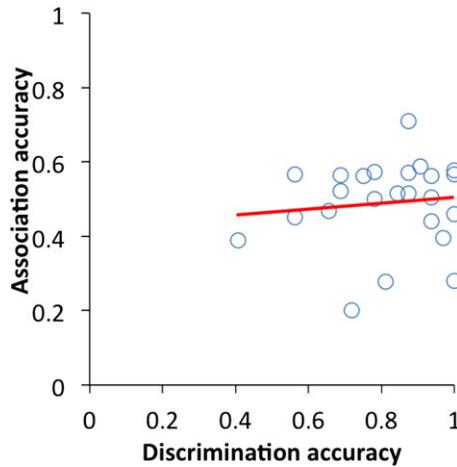
Fig. 4. Experiment 2, discrimination accuracy versus association accuracy. FA = false alarms.

eight critical-different trials (dropping the four trained-different trials), and all eight of the same trials (four critical-same and four trained-same), so that biases to say "same" or "different" more often would be canceled out. Overall, children showed above-chance discrimination accuracy ($81.5 \pm 16.1\%$; $t(23) = 9.52$, $p < .0001$, Cohen's $d = 1.94$).

3.2.2. Association

Despite good discrimination performance, children were overall at chance on association ($M = 47.7\%$, $SD = 11.9\%$; $t(23) = 1.03$, $p = .32$, Cohen's $d = .21$). However, it was possible that learning performance was linked to discrimination performance. If so, then discrimination accuracy should correlate positively with association accuracy. However, this correlation (Fig. 4) did not approach significance ($r(22) = .04$, $p = .87$; note that the slope in Fig. 4 appears slightly more positive because it was calculated based on raw values, whereas the slope reported in the text used e-logit transformed values). No correlations between age and either dependent measure approached significance.

3.2.3. Effects of tone language

One might ask whether children with tone-language exposure would perform better on either task, given demonstrations that adult tone-language speakers perform better on relative-pitch tasks (Pfordresher & Brown, 2009). While the current experiment was not designed to test this hypothesis, it is important to verify that good discrimination results were not driven by the tone-language-speaking children alone. Exploratory analyses revealed no effects of tone language exposure. Both groups exceeded chance on the discrimination task ($p \leq .002$) and did not differ (tone: 82.6% vs. non-tone: 79.3%; Welch's t for unequal variances: $t(13.3) = 0.09$, $p = .93$, Cohen's $d = .04$). Neither group exceeded chance on the association task, and again there was no difference between groups (50.0% vs. 43%; Welch's t for unequal variances: $t(13.4) = 1.31$, $p = .21$, Cohen's

d = .58). While this does not imply the absence of group differences in childhood pitch processing based on tone language exposure, it does suggest that such effects, if found, would be modest in magnitude.

3.3. Discussion

This experiment replicated Experiment 1 in showing that preschoolers have difficulty associating melodies distinguished by pitch contour to visual referents. Furthermore, it suggests that preschool children *discriminate* different-contour melodies relatively well in an explicit task, consistent with a more mature, explicit version of infants' presumably implicit contour discrimination abilities (Trehub et al., 1984). This suggests that immediate representations of contour may precede stable long-term representations of contour, like those needed for the association task. Alternately, it may suggest that in an easier task (immediate discrimination), relatively weak representations of the melodies suffice, whereas in a harder task (association), weak representations are insufficient.

There is one alternative interpretation of the results that should be addressed here. Namely, it is logically possible that children tested here are overall very poor at pitch contour and succeed in the discrimination task using differences in rhythm (see Fig. 1), not contour. It is also possible that tone language experience contributed to unusually good discrimination performance (though exploratory comparisons suggested this was not the case). To assess this, Experiment 2 was replicated (henceforth referred to as Experiment 2B) with an additional 24 children (none of whom were tone language speakers) using melodies differing only in contour (F#4-G#4-A#4-B4-C#5 vs. C#5-B4-A#4-G#4-F#4). Children readily discriminated melodies differing only in contour ($M = 69.8 \pm 21.8\%$; $t(23) = 4.58$, $p = .0001$, Cohen's $d = 0.93$).[3] As stated earlier, the audiovisual association test showed chance performance ($M = 48.2 \pm 22.0\%$, $t(23) = -0.09$, $p = .93$, Cohen's $d = -0.02$). This occurred even though mapping test instructions were altered to ask children "Whose favorite song was that?" rather than asking who was singing their favorite song, implying that the particular instructions used did not strongly affect the mapping outcome. This suggests that, even in the absence of the rhythmic difference, children discriminated the melodies but did not maintain audiovisual associations.

The final experiment tested children's sensitivity to a different cue: timbre. This allowed assessment of timbre sensitivity in children, previously studied mainly in adults (e.g., Halpern & Müllensiefen, 2007). It also allowed a counter to the possibility that the association task itself is too difficult for children (though note that this paradigm has previously been used successfully in word–picture associations [Creel, 2014a, b] and voice–picture associations [Creel & Jimenez, 2012]). If the association task in Experiments 1 and 2 was simply globally "too hard," then children in Experiment 3 should perform similar to those in Experiment 2: They should succeed in discrimination but fail in the association task. These predictions also hold if presenting the association task after another experimental phase (listening or discrimination) is too taxing for children. However, if children in Experiments 1 and 2 failed in the association task because contour

representations are weaker or less durable in memory than timbre is, then children in Experiment 3 should perform well in both discrimination and association.

4. Experiment 3

4.1. Method

4.1.1. Participants

Twenty-four new 3–5-year-old children (13 female; $M = 4.8$, $SD = 0.6$) recruited from schools similar to those in previous experiments took part. Twenty of the children heard other languages in addition to English, including six children who heard tone languages (3 Korean, 2 Mandarin, 1 Vietnamese). Five more children were tested but not included due to shyness (2), computer error (2), or interruption (1).

4.1.2. Stimuli

Stimuli were similar to those in Experiment 2, except that the different-contour stimuli during discrimination, audiovisual training, and audiovisual test were replaced with different-*timbre* stimuli (rising melody played by muted trumpet, rising melody played by vibraphone). These timbres were chosen because Iverson and Krumhansl (1993) reported that adults find them highly distinct.

4.1.3. Procedure

This was identical to Experiment 2.

4.2. Results

4.2.1. Discrimination

Most children (22) passed the reinforced training in one 8-trial block; one needed two blocks; one timed out after three blocks. As stated earlier, all children were included for maximum comparability across experiments. Children showed 95.4% accuracy ($SD = 11.8\%$), which exceeded chance performance ($t(23) = 17.61$, $p < .0001$, Cohen's $d = 3.59$). There was a trend for higher accuracy for older children ($r(22) = .39$, $p = .07$), but this was driven by the youngest participant, who was the only one who failed to reach criterion performance on the reinforced training phase. Discrimination was better than in Experiment 2 ($t(46) = 3.77$, $p = .0005$, Cohen's $d = 1.09$).

4.2.2. Association

Children were highly accurate ($M = 89.1\%$, $SD = 16.0\%$; Fig. 3, right), performing well above chance ($t(23) = 9.39$, $p < .0001$, Cohen's $d = 1.92$). There was a trend for higher accuracy in older children ($r(22) = .38$, $p = .07$, though the younger half of children still exceeded chance accuracy). Association accuracy was higher in the current experiment than in Experiment 2 ($t(46) = 9.15$, $p < .0001$, Cohen's $d = 2.64$). However,

as discrimination accuracy was also better in the current experiment, the comparison was tested with discrimination accuracy partialed out. Experiment 3 still showed higher accuracy than Experiment 2 ($F(1,45) = 56.94$, $p < .0001$, $\eta^2_P = .56$).

4.2.3. Effects of tone language

As in Experiment 2, exploratory analyses examined effects of tone-language exposure on performance. Like Experiment 2, no effects of tone language emerged. Both groups exceeded chance on the discrimination task (tone: 95.8%, non-tone: $M = 95.3\%$), and the two groups did not differ (Welch's t for unequal variances: $t(16.3) = 0.10$, $p = .92$, Cohen's $d = .04$). Both groups also exceeded chance on the association task (94.5% vs. 87.2%) and did not differ from each other (Welch's t for unequal variances: $t(11.3) = 1.11$, $p = .29$, Cohen's $d = .45$).

4.3. Discussion

In this experiment, association accuracy was high even though the learning task was formally identical to that in the first two experiments. This suggests that, for children, timbre information may be more robust than pitch contour in associative learning. This result holds even when discrimination accuracy is partialed out. The reasons why timbre might be more robust, and the implications of this finding for developing auditory representations, are considered below.

5. General discussion

The study presented here aimed to assess what types of information preschool-aged children store in memory about the music they hear, specifically, whether they privilege musically relevant aspects (pitch contour) over less-relevant ones (timbre). Association learning was assessed by means of a sound–picture association task, where children were asked to recognize which of two cartoon characters' "favorite song" was playing. Discrimination ability was assessed in an immediate same-different task. Across three experiments, children showed good same-different discrimination for both pitch and timbre, but they showed sound–picture association only for timbre. Thus, while the discrimination task reflects sensitivity of both timbre and pitch contour, children's association learning preserves timbre more strongly than contour. A later section of the General Discussion explores why this might be the case. In any event, this outcome suggests that the answer to the question raised in the Introduction—whether children aged 4–5 years have learned which aspects of music are relevant and which are not—is either "no" or "not completely."

The study additionally asked whether exposure aided in formation of representations. While this study cannot rule out effects of exposure, the magnitude of exposure effects in the current age group was so small as to be unobservable. Experiment 1 manipulated the type of exposure (variable, invariable, or none), and found no differences between conditions. Experiments 2 and 3 exposed children to melodies in the context of a

discrimination task. One might think that focusing children on differences between the melodies would improve performance in the association task. However, Experiment 2, which tested pitch contour memory, showed performance at chance, equivalent to Experiment 1, which gave children only incidental exposure. Nevertheless, previous work suggests that much lengthier exposure (hours, days, or longer) does generate stable memory associations for musical attributes that are fairly generalizable over a change in timbre (Vongpaisal et al., 2009), and that those representations include contour and exact pitch intervals (Corrigall & Trainor, 2010). This implies that the amount or type of exposure presented here is insufficient to generate memories durable enough to allow association formation.

5.1. Various interpretations

There are at least three interpretations of this set of results. Two interpretations concern memory representations. First, the results may reflect a difference in *representation strength*. Second, there may be a distinction in *memory duration*, with contour memory but not timbre memory being difficult to preserve for long durations. On a representation strength account, contour may be represented more weakly in memory from the outset. That is, perhaps timbre is simply more salient and memorable to children of this age than are pitch contour patterns (see Curtin, Fennell, & Escudero, 2009, for a similar result in language where children distinguish some vowel-differing words earlier than children in Stager & Werker, 1997, distinguished consonant-differing words). This would mean that there are initial differences in encoding of contour versus timbre information, such that discrimination performance as well as mapping would be weaker for contour than for timbre—and indeed discrimination performance was weaker for contour (Experiment 2) than for timbre (Experiment 3), though both exceeded chance (see Experiment 3 Results section). Of course, even considering the children who scored perfectly or near-perfectly on the discrimination task in Experiment 2 (rightmost points in Fig. 4), association was still at chance. Thus, if salience or strong discriminability constrains association learning, it seems likely that the salience threshold for mapping is far above the level that permits discrimination.

On a memory *duration* account, contour might initially be represented accurately, but fade more rapidly over time than timbre memory. If so, there would be no reason to predict initial weaknesses for contour encoding, just a greater decline over time. A third possibility is that the tasks themselves might be qualitatively different, such that contour matters less for the association task than for the discrimination task. Perhaps the association task places additional demands on children to form and remember *associations*, not just remember the melodies themselves. While this is possible, it seems logical that if association formation were the most difficult element of the task, children would have failed in association formation across the board. Yet they were near ceiling accuracy in Experiment 3. Nonetheless, it is interesting to consider whether a contour discrimination task with a long memory duration would pattern differently than association learning. A memory duration account suggests that children would do very poorly in both association

and in discrimination tasks with long delays. An account citing association learning as the predominant difficulty suggests that children would fail at association but would do well in a long-delay discrimination task (where no association is required). In any case, though, there needs to be some way to account for the stronger performance on timbre association learning.

5.2. Why would timbre be encoded more readily than contour?

One likely candidate explanation for ready encoding of timbre but not contour is that children may generally pay little attention to pitch contour information, perhaps shaped by language experience (though note that Fernald's work [1989; and Fernald & Kuhl, 1987] suggests that pitch characteristics of infant-directed speech drive infant attention very early in life). Of course, evidence from infants (Mattock & Burnham, 2006) and adults (Burnham et al., 1996) suggests that pitch discrimination in music is preserved even as pitch contrasts in language become less salient to non-tone language speakers. Perhaps, then, pitch is not unlearned but is simply more difficult to encode than timbre. Some research on linguistic and paralinguistic processing supports this. Children learning non-tone languages do not use speech pitch content for emotion detection until age 4 or 5 (Quam & Swingley, 2012), well after they use faces and body postures to detect emotion (Nelson & Russell, 2011). Children growing up learning tone languages show maintained early discrimination of pitch contrasts in language (Mattock & Burnham, 2006; Singh & Foong, 2012), yet they are less accurate than adults in tone perception or production at age 3 in Mandarin (Wong, Schwartz, & Jenkins, 2005) and may not be adult-like in their tone recognition until age 10 in Cantonese (Ciocca & Lui, 2003). These outcomes suggest that full adult-like pitch contour representations, despite evidence of early infant sensitivity to exaggerated pitch contours (e.g., Fernald, 1989), may require a lengthy learning process. Note that this does not imply that pitch is "acquired" wholesale at a particular age in development, but that children must experience *particular* pitch patterns multitudinously to form robust representations of them.

Another explanation of better timbre encoding than pitch encoding is that events with changing elements (pitches) are harder to encode than events with unchanging elements (timbres). One might quantify changeability by stating that contour or scale degree has a higher *information content* (Pearce & Wiggins, 2006)—that is, lower predictability—than timbre does. In real music, like the melodies used here, pitch contour changes much more often than timbre does—there are many more possibilities for sequential combinations of contour or scale degree than there are distinct timbres. Thus, the stimuli used here are reflective of the statistics of real music. Nonetheless, the information content hypothesis could be further explored with variable-timbre sequences. If children are simply better at encoding spectral content than pitch content, variable-timbre sequences should be easy to encode. If children specifically have difficulty with high information content or order, then variable-timbre sequences should be hard to encode, as they would have a high information content as well. An alternative possibility is that very high information content would be more likely to elicit attention (e.g., longer and more complex melodies),

predicting that children might be *more* drawn to differences between complex melodies and encode them better than the simple melodies used here.

A third possible explanation for why timbre would be more easily associated with objects than contour relates to the role of timbre in sound source identification. That is, perhaps a timbre is easier to map to an object because timbre, but not melodic contour, tells one something about an object's physical shape as a resonating body. Relatedly, a particular timbre, but not a particular melody, readily picks out a particular object—a musical instrument. For instance, a harp can play many melodies but has a distinct timbre regardless.[4] Thus, on both these accounts, timbres might be more likely to map to object identity than melodies are. In this vein, one might ask whether instructing children to point to the character who is *singing* its favorite song might bias them to attend to character-inherent properties such as resonance or voice quality—to timbre, facilitating timbre mapping and depressing pitch mapping performance. While this is possible, two things suggest that it is unlikely. First, the sound quality of the musical instruments was not voice like. More important, a change in instructions in Experiment 2B—asking "whose favorite song is it" rather than "which character is singing"—did not improve accuracy over Experiment 2.

5.3. Limitations and future directions

While this study suggests differences in children's sensitivity to timbre versus contour, there are still a number of open questions. One question is whether the familiarization was simply insufficient to yield good representations of contour. At one level this is trivially true: Work by Corrigall and Trainor (2010), and the popularity of children's choirs, attest that children can form robust representations of pitch contour given sufficient exposure. However, the amount of this exposure may be quite large. In this study, children heard a maximum of 44 repetitions per 2-s melody (36 times in Experiment 1's preexposure, plus eight times during the learning phase), totaling roughly 90 s. In contrast, children in the same age group in Vongpaisal et al. (2009) had likely heard numerous repetitions of cartoon TV show themes: Assuming conservatively that each theme was 1 min long, and that the child had seen the show 30 times, this constituted at least 30 min of exposure. Furthermore, the exposure was distributed over multiple days, providing not only distributed learning (e.g., Cepeda et al., 2009) but also ample time for sleep consolidation processes (e.g., Stickgold & Walker, 2005). It may be that both lengthy exposure and sleep consolidation are required for forming durable representations of contour, perhaps due to its high information content. The implication across studies, perhaps, is that particular pitch patterns—whether they constitute melodies, signal vocal affect, or indicate question versus statement intonation—are learned over a lengthy time course and across many instances.

A related limitation is that the familiarization was timbre-invariant (though it was pitch-variable for some children in Experiment 1). Perhaps children need to hear melodies in a range of timbres, just as they may need to hear words in a range of voices (Rost & McMurray, 2009, 2010), in order for the melodic contour to emerge as the consistent

characteristic. Nonetheless, timbre invariance is a reality for many modern musical experiences: Pop songs (see Levitin, 1994), ubiquitous cell phone ringtones, and even orchestral recordings rarely change timbre from hearing to hearing. Other types of music, particularly songs sung with children or at social occasions, may exhibit more timbre and pitch variability (see, e.g., Halpern, 1989; though see Bergeson & Trehub, 2002, for data on pitch constancy in infant-directed singing). Thus, the exposure provided here may be representative of only certain types of music experienced in daily life, urging exploration of variable-timbre exposure.

A third limitation is that the contour difference used here may have been too subtle for child participants. While the two melodies differed completely in pitch direction, the intervals between successive notes were modest (see Stalinski et al., 2008, on children's sensitivity to pitch direction at different interval sizes). This raises the question of how to define a sufficiently "strong" difference in pitch contour. One reasonable criterion might be the frequency with which certain contours are encountered in real music. Most note-to-note transitions in Western music differ by two semitones or less (the average note-to-note pitch distance in the contour-differing melodies here; Watt, 1924; see also Dowling, 1978). This suggests that the contour distinction and interval sizes used here were reasonably reflective of real-world musical exposure. While one could use larger step sizes, such large intervals are not typical of musical experience. Nonetheless, an additional experiment presented children with the association task only, using 5-note melodies that rose or fell over a 2-octave range (24 semitones)—more than three times the pitch range (7 semitones) of the contours in Experiments 1 and 2. This experiment found no evidence of learning ($M = 52\% \pm 26\%$, $t(22) = 0.76$, $p = .45$). Thus, increasing the starkness of the contour changes did not increase association learning.[5]

An interesting possibility relates to task pragmatics: Might children perform better if the referential nature of the melodies were made clear? In the Switch paradigm with infants, Fennell and Waxman (2010) have found that younger infants (14 months) succeed in mapping similar-sounding words to objects when they hear words embedded in sentences, rather than the isolated word tokens typically heard in Switch experiments. Even more interesting, Campbell and Namy (2013) found that 13- and 18-month-olds can learn that non-speech sounds refer to objects, but only when the pragmatics strongly suggest that objects are being labeled (e.g., "Do you see what we have there? <BEEP-BOOP>!"). An additional point with regard to pragmatics is that the use of animate cartoons in the current task may have implied that the labels were proper names rather than common nouns labeling categories, which differs from many previous word-learning studies. Whether a proper-noun interpretation might impede learning is unclear.

Finally, one might wonder whether the sound–picture association task underestimates long-term memory storage because it does not map on to the types of melody-referent associations children are accustomed to making. For example, children may expect music to cross-modally match the character in some way. For instance, perhaps a bouncy or spiky creature has a staccato melody (short, separated notes), whereas a more fluidly moving or bulbous creature has a more legato melody (smooth, continuous notes). In the current case, rising creatures might associate more easily with rising melodies, and falling

creatures with falling melodies. Future studies could investigate whether such cross-modal, non-arbitrary mappings might generate better performance (see Walker et al., 2010, for some evidence consistent with this idea in young infants).

5.4. Developing auditory perception

What do these results suggest about auditory development? One implication is that good discrimination does not reflect strong or stable representations. Thus, earlier studies showing contour discrimination in much younger children (Chang & Trehub, 1977; Trehub et al., 1984) may not constitute evidence of stable memory representations themselves, but the *capacity to form* such representations. This is in no way meant to suggest that infants and preschoolers do not *possess* representations of speech sounds or pitch contours—they undoubtedly do—but that the representations they use in discrimination tasks may be relatively weak traces, or that their representations are so implicit that they are not evident in a behavioral task (e.g., Corrigall & Trainor, 2014). Also of interest, this resembles the pattern that Stager and Werker (1997, Werker et al., 2002) observed in much younger children (14 months): good discrimination of similar-sounding words, but no evidence of mapping those words to different pictures. Of course, given the differences in age (infant vs. 4–5 years), experimental paradigm, and domain (language vs. music), one might reasonably question whether Stager and Werker's findings and the current findings exemplify the same phenomenon. If they do, one tentative interpretation is that tasks which demonstrate infants' or children's ability to discriminate stimuli in music, speech, and other domains reflect the beginnings of a representational system rather than a mature one.

How are children's representations of sounds organized? One possibility is that early sound representations encode contour-only weakly, perhaps because of its high information content (unpredictability) relative to more static musical features. Contour becomes more salient, and gains greater perceptual weight, as learners slowly discern that pitch contour (and relative pitch) is a criterial cue to melodic identity. This account also fits with early perception but late comprehension of pitch in speech prosody: While infants respond differently to different vocal emotions (e.g., Singh, Morgan, & Best, 2002), they have difficulty *linking* pitch patterns to a speaker's emotions until about 4 years (Quam & Swingley, 2012) or even later (see Morton & Trehub, 2001). That is, on this account, children take a long time to process melodic contour or to map prosodic patterns to affective states because it takes children a long time to learn that pitch patterns are important. Of course, this account would suggest that pitch processing should either become accurate in music and prosody at around the same developmental time or that pitch processing develops separately in each domain.

A slightly different account of the development of contour perception is that children begin with no dimensions at all, but initially organize memory around highly familiar instances. Only for highly familiar instances is processing highly accurate. Dimensions emerge only slowly as more information is accrued, with initial salience governing memory robustness, whereas less salient but more diagnostic dimensions like contour emerge later. Organization of memory around highly familiar instances is supported by evidence

of greater 4-year-old musical sensitivities in familiar melodies (Corrigall & Trainor, 2010) than unfamiliar melodies (Trainor & Trehub, 1994). This also fits with word-recognition data from Swingley and Aslin (2002), who showed that 14-month-olds were sensitive to subtle mispronunciations of familiar words, the same age where Stager and Werker's (1997) children were insensitive to subtle differences in *novel* words. This account provides a more realistic mechanism than the previous account for explaining why pitch processing might appear to develop differently in different domains: Children process *particular* pitch patterns well as a function of massive exposure, rather than improving in processing pitch as a unified dimension.

6. Conclusion

This study implies that 4–5-year-olds represent timbre more robustly than pitch contour, even though pitch contour is an important cue to musical identity. The results differ from numerous previous demonstrations of infant *sensitivity* to pitch contour in immediate memory (e.g., Trehub et al., 1984), possibly because of greater sensitivity in the infant paradigms used. Results are consistent with preserved memory for timbre in adult listeners (Halpern & Müllensiefen, 2007; Radvansky & Potter, 2000; Radvansky et al., 1995). More broadly, the current results suggest that children's memory representations of non-speech auditory materials are still under development in the preschool years and may be less advanced than their spoken language representations, which are already fairly robust to irrelevant variability.

Notes

1. A phonemic change is a linguistic sound change that can distinguish meaning in a particular language, such as the change from /g/ to /l/ in the dog–doll example above.
2. This distribution of Vietnamese speakers is due to two preschools with high concentrations of Vietnamese speakers being scheduled when this experiment was being run.
3. An additional discrimination experiment from a related but separate study included a contour-only condition. Supporting the current results, children in that experiment

also discriminated contour-differing melodies, $t(47) = 5.49$, $p < .0001$, Cohen's $d = 0.76$.

4. It is tempting to offer voices as an example of children's abilities to map timbres to visual objects (faces), but previous research indicates that children at and above the current age range are actually fairly poor at identifying voices (Bartholomeus, 1973; Creel & Jimenez, 2012; Mann, Diamond, & Carey, 1979), and are particularly poor at voice-character mapping in a paradigm nearly identical to this study, if voices are highly similar; they do well when they are learning different-age or different-gender voices (Creel & Jimenez, 2012). Furthermore, early work demonstrated that there is not a uniform "timbre" that characterizes an individual's voice (Bricker & Pruzansky, 1966), because speech sound quality (aa, oo, ee) also affects vocal timbre.

5. Interestingly, in this experiment, in Experiment 2B, and in a handful of children in Experiment 2, children showed somewhat consistent trial-to-trial responding (see Appendix B on locally consistent responding). That is, despite chance *accuracy*, their response on trial *n* was much more likely than chance to reflect the same association as trial *n* − 1. This suggests that they could maintain an association pattern for the duration of one or more trials, despite not being able to maintain associations for the longer time interval from training to test. This is also consistent with the brief-duration association formation observed by Horst and Samuelson (2008) in word learning.

References

Backscheider, A. G., & Gelman, S. A. (1995). Children's understanding of homonyms. *Journal of Child Language*, *22*, 107–127.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology*, *27*(4), 464–472. Available at http://www.ncbi.nlm.nih.gov/pubmed/4766153. Accessed August 4, 2010.

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11892783. Accessed October 9, 2013.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America*, *40*(6), 1441–1449. doi: 10.1121/1.1910246.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Lacerda, F., & Attapaiboon, C. (1996). Facilitation or attenuation in the development of speech mode processing? Tone perception over linguistic contexts. In P. McCormack & A. Russell (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 587–592). Canberra, Australia: Australian Speech Science and Technology Association.

Campbell, A. L., & Namy, L. L. (2013). The role of social-referential context in verbal and nonverbal symbol learning. *Child Development*, *74*(2), 549–563. Available at http://www.ncbi.nlm.nih.gov/pubmed/12705572. Accessed May 5, 2013.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. In *Papers and reports on child language development 15* (pp. 17–29).

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. doi: 10.1027/1618-3169.56.4.236.

Chang, H.-W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, *24*, 324–331.

Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*, *1*(2), 141–147. doi: 10.1080/1476967031000090971.

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, *28*(2), 195–200.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142–158. doi: 10.1111/desc.12100.

Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development*, *83* (6), 2042–2056. doi: 10.1111/j.1467-8624.2012.01816.x.

Creel, S. C. (2014a). Impossible to ignore: Word-form inconsistency slows preschool children's word-learning. *Language Learning and Development*, *10*(1), 68–95. doi: 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.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*, 633–664. doi: 10.1016/j.cognition.2007.03.013.

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509.

Curtin, S., Fennel, C. & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–31. doi: 10.1111/j.1467-7687.2009.00814.x.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10. doi: 10.1016/S0010-0277(00)00136-0.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422. doi: 10.1016/j.cognition.2009.10.013.

Fennell, C., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*(5), 1376–1383. doi: 10.1111/j.1467-8624.2010.01479.x.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. doi: 10.1016/0163-6383(87)90017-8.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D. & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573–576. doi: 10.1016/j.cub.2009.02.058

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(5), 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, *105* (2), 251–279.

Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology*, *35*(1), 294–302. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9923483. Accessed September 4, 2013.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, *17*(5), 572–581. Available at http://www.ncbi.nlm.nih.gov/pubmed/2796742. Accessed September 1, 2009.

Halpern, A. R., & Müllensiefen, D. (2007). Effects of timbre and tempo change on memory for music. *Quarterly Journal of Experimental Psychology*, *61*(9), 1371–1384. doi: 10.1080/17470210701508038.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, *50*, 354–377. doi: 10.1016/j.cogpsych.2004.09.003.

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48–55.

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639–12643.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. doi: 10.1080/1525O(xH)701795598.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. doi: 10.10371/0096-1523.26.5.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, *94*(5), 2595–2603. Available at http://www.ncbi.nlm.nih.gov/pubmed/8270737. Accessed October 5, 2010.

Keller, T. A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Developmental Psychology*, *30*(6), 855–863. doi: 10.1037//0012-1649.30.6.855.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423.

Locke, R. P. (1986). Program music. In D. M. Randel (Ed.), *New Harvard dictionary of music* (pp. 656–659). Cambridge, MA: Harvard University Press.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*(4), 272–276.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah, NJ: Erlbaum.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*, 153–165.

Massaro, D. W., Kallman, H. J., & Kelly, J. L. (1980). The role of tone height, melodic contour, and tone chroma in melody recognition. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(1), 77–90.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. doi: 10.1111/j.1467-7687.2007.00653.x.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834–843.

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52–61. doi: 10.1016/j.jecp.2011.03.014.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*(2), 309–328.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377–405.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*(6), 1385–1398. doi: 10.3758/APP.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1–11. doi: 10.1016/j.cognition.2004.09.008.

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250. doi: 10.1111/j.1467-8624.2011.01700.x.

Radvansky, G. A., Fleming, K. J., & Simmons, J. A. (1995). Timbre reliance in nonmusicians' and musicians' memory for melodies. *Music Perception*, *13*(2), 127–140.

Radvansky, G. A., & Potter, J. K. (2000). Source cuing: Memory for melodies. *Memory & Cognition*, *28*(5), 693–699. Available at http://www.ncbi.nlm.nih.gov/pubmed/10983442. Accessed August 12, 2013..

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

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. doi: 10.1111/j.1532-7078.2010.00033.x.

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: An event-related potential study with ringtones. *European Journal of Neuroscience*, *26*(3), 784–790. doi: 10.1111/j.1460-9568.2007.05685.x.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*(2), 113–129. doi: 10.1016/j.cognition.2005.01.004.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R., Loman, M. M., & Robertson, R. R. (2000). Infant memory for musical experiences. *Cognition*, *77*(1), B15–B23. Available at http://www.ncbi.nlm.nih.gov/pubmed/10980255. Accessed October 1, 2009.

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *6*(4), 641–646. Available at http://www.ncbi.nlm.nih.gov/pubmed/10682207. Accessed September 25, 2006.

Schellenberg, E. G., & Trehub, S. E. (1996). Children's discrimination of melodic intervals. *Developmental Psychology*, *32*(6), 1039–1050.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*(3), 262–6. Available at http://www.ncbi.nlm.nih.gov/pubmed/12741751 Accessed May 3, 2004.

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, *124*(2), 128–142. doi: 10.1016/j.cognition.2012.05.008.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365–394. doi: 10.1207/S15327078IN0303_5.

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.

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: Distinguishing up from down. *Journal of the Acoustical Society of America*, *124*(3), 1759–1763. doi: 10.1121/1.2956470.

Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences*, *28*(8), 408–415. doi: 10.1016/j.tins.2005.06.004.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*, 480–484.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*, 16–34. doi: 10.1016/j.jml.2006.07.002.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*(2), 125–132. Available at http://www.ncbi.nlm.nih.gov/pubmed/7971113. Accessed October 11, 2009.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*(3), 289–296. Available at http://www.ncbi.nlm.nih.gov/pubmed/15595370. Accessed October 25, 2005.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 295–301. Available at http://www.ncbi.nlm.nih.gov/pubmed/2943857. Accessed September 11, 2011.

Trehub, S. E., Endman, W., & Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, *313*, 300–313.

Trehub, S. E., Schellenberg, E. G., & Nakata, T. (2008). Cross-cultural perspectives on pitch memory. *Journal of Experimental Child Psychology*, *100*(1), 40–52. doi: 10.1016/j.jecp.2008.01.007.

Vongpaisal, T., Trehub, S. E., & Schellenberg, E. G. (2009). Identification of TV tunes by children with cochlear implants. *Music Perception*, *27*(1), 17–24.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–25. doi: 10.1177/0956797609354734.

Watt, H. J. (1924). Functions of the size of interval in the songs of Schubert and of the Chippewa and Teton Sioux Indians. *British Journal of Psychology*, *14*(4), 370–386.

Werker, J. F., Fennell, C., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. doi: 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*(1), 49–63. doi: 10.1016/S0163-6383(84)80022-3.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, *48*, 1065–1079.

Yoshida, K. A., Fennell, C., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–418. doi: 10.1111/j.1467-7687.2008.00789.x.

Appendix A: Effects of data transformations on results

Discrimination scores were calculated based on the eight critical-different trials and all eight same trials (four critical-same and four trained-same). For raw and empirical-logit scores, accuracies were averaged. For d-prime scores, hits (saying "different" on different trials) and false alarms (saying "different" on same trials) were incorporated.

For d-prime on association accuracy, the Melody 1-Picture 1 mapping was arbitrarily defined as the "target." Thus, a "hit" was defined as selecting Picture 1 when hearing Melody 1. A "false alarm" was correspondingly defined as selecting Picture 1 when hearing Melody 2. This is mathematically identical to defining Picture 2 as the target, where a hit is selecting Picture 2 for Melody 2, and a false alarm is selecting Picture 2 for Melody 1. Note that either formulation takes all trials into account and is highly correlated with raw accuracy.

For raw scores, chance = .50; for e-logit and d-prime, chance = 0.0. D-prime for the yes-no task is simply z(Hits)—z(False Alarms). D-prime for the two-alternative association task is the same except that it is divided by the square root of 2 (see MacMillan & Creelman, 2005, Ch. 7).

Bolded comparisons are statistically significant; grayed-out comparisons are not statistically significant.

Experiment 1

	M accuracy	Conditions	Accuracy versus Chance	No Exposure versus Chance	Same-pitch Exposure versus Chance	Var.-pitch Exposure versus Chance
raw	.50 (.14)	$F_{(2,21)} = .61, p = .55$	$t_{(23)} = 0.09, p = .93$	$t_{(7)} = 1.27, p = .24$	$t_{(7)} = -0.35, p = .74$	$t_{(7)} = -0.42, p = .68$
e-logit	.00 (.57)	$F_{(2,21)} = .58, p = .57$	$t_{(23)} = -0.01, p = .99$	$t_{(7)} = 1.23, p = .26$	$t_{(7)} = -0.42, p = .69$	$t_{(7)} = -0.40, p = .70$
d-prime	0.00 (0.54)	$F_{(2,21)} = .58, p = .57$	$t_{(23)} = 0.00, p = 1.00$	$t_{(7)} = 1.23, p = .26$	$t_{(7)} = -0.41, p = .70$	$t_{(7)} = -0.41, p = .69$
binomial	11/5/8[a]	—	$p = .65$	$p = .13$	$p = 1.00$	$p = .69$

[a]Successes (>.50)/failures (<.50)/chance (=.50). Scores exactly at chance were not entered into the binomial test.

Experiment 2

	M Accuracy	M Discrimination	Accuracy versus Chance	Discrimination versus Chance	Accuracy-Discrimination Correlation
raw	.48 (.12)	.82 (.16)	$t_{(23)} = -0.96, p = .35$	$t_{(23)} = 9.61, p < .0001$	$r_{(22)} = .10, p = .64$
e-logit	-0.10 (0.48)	1.65 (0.85)	$t_{(23)} = -1.03, p = .32$	$t_{(23)} = 9.52, p < .0001$	$r_{(22)} = .04, p = .87$
d-prime	-0.09 (0.47)	1.95 (0.99)	$t_{(23)} = -0.97, p = .34$	$t_{(23)} = 9.63, p < .0001$	$r_{(22)} = .03, p = .90$
binomial	10/9/5	23/1/0	$p = .53$	$p < .0001$	—

Experiment 2B

	M Accuracy	M Discrimination	Accuracy versus Chance	Discrimination versus Chance	Accuracy-Discrimination Correlation
raw	.48 (.22)	.70 (.22)	$t_{(23)} = -0.41, p = .69$	$t_{(23)} = 4.45, p = .0002$	$r_{(22)} = .09, p = .67$
e-logit	-0.02 (1.10)	1.09 (1.17)	$t_{(23)} = -0.09, p = .93$	$t_{(23)} = 4.58, p < .0001$	$r_{(22)} = .04, p = .84$
d-prime	-0.05 (0.90)	1.21 (1.31)	$t_{(23)} = -0.27, p = .79$	$t_{(23)} = 4.53, p = .0001$	$r_{(22)} = .05, p = .81$
binomial	10/12/2	17/3/4	$p = .83$	$p = .003$	—

Experiment 3

	M Accuracy	M Discrimination	Accuracy versus Chance	Discrimination versus Chance	Accuracy-Discrimination Correlation
raw	.89 (.16)	.95 (.12)	$t_{(23)} = 11.97, p < .0001$	$t_{(23)} = 18.92, p < .0001$	$r_{(22)} = .24, p = .25$
e-logit	2.45 (1.28)	2.50 (0.70)	$t_{(23)} = 9.39, p < .0001$	$t_{(23)} = 17.61, p < .0001$	$r_{(22)} = .20, p = .34$
d-prime	1.68 (0.71)	2.75 (0.75)	$t_{(23)} = 11.62, p < .0001$	$t_{(23)} = 18.06, p < .0001$	$r_{(22)} = .28, p = .18$
binomial	23/1/0	23/1/0	$p < .0001$	$p < .0001$	—

Experiment 2 versus Experiment 3

	Accuracy Difference	Discrimination Difference	Accuracy Controlled for Discrimination
raw	$t(46) = 10.17, p < .0001$	$t(46) = 3.43, p = .002$	$F(1,45) = 73.97, p < .0001$
e-logit	$t(46) = 9.15, p < .0001$	$t(46) = 3.77, p = .0005$	$F(1,45) = 56.94, p < .0001$
d-prime	$t(46) = 10.21, p < .0001$	$t(46) = 3.16, p = .003$	$F(1,45) = 77.82, p < .0001$
Fisher's exact test	$p = .002$	$p = 1.00$	–

Appendix B: Response patterns

Shown here are participants who showed above-chance response consistency (roughly 28% of children in Experiments 1, 2, and 2B). This meant that 13 or more of 16 total responses conformed to some pattern ($p < .05$ by two-tailed binomial test). Combined across Experiments 1, 2, and 2B, consistent responders were slightly but significantly older ($M = 4.91$ years, $SD = .55$) than non-consistent responders ($M = 4.59$, $SD = .62$; $t(38.6) = 2.14, p = .04$).

A small number of children showed what might be considered visual responses. Some tended to select the object on one side of the screen, whereas a few others tended to select the same picture on every trial. Others responded in a manner that appeared to be contingent on sound-object mappings. Some responded correctly on many trials, but just as many responded *incorrectly* on just as many trials. This suggests that perhaps those children are picking a melody-to-creature-mapping at the start of the test, and then continuing to adhere to that mapping. Furthermore, some children appeared use the same mapping for several trials in a row, occasionally switching to the opposite mapping. Specifically, their response indicated a melody-to-character association that was consistent with the association on the previous trial (Table B1). This locally consistent mapping pattern could only be assessed for trials 2–16, as the first trial had no preceding trial. Responding based on the same mapping as the previous trial for 12 or more of 15 trials is less probable than .05 (two-tailed binomial test).

Note that near-100% and near-0% accuracy also register as highly consistent, but participants are only reported as locally consistent if they were not already reported as using a correct mapping or a reversed mapping. Importantly, consistent responses (fully correct, fully incorrect) can be made based simply on the memory of the previous trial. Overall, these results suggest that some children attempted to use a response rule but may not have had access to the mappings demonstrated to them during the learning phase.

Table B1
Example of coding local consistency

Trial	Melody	Chose Picture	Accuracy	Consistency	Mapping Used
1	Mel2	Pic2	1	(first trial)	1 = 1, 2 = 2
2	Mel2	Pic2	1	consistent	1 = 1, 2 = 2
3	Mel1	Pic1	1	consistent	1 = 1, 2 = 2
4	Mel1	Pic1	1	consistent	1 = 1, 2 = 2
5	Mel2	Pic1	0	**inconsistent**	**1 = 2, 2 = 1**
6	Mel1	Pic2	0	consistent	1 = 2, 2 = 1
7	Mel2	Pic1	0	consistent	1 = 2, 2 = 1
8	Mel1	Pic2	0	consistent	1 = 2, 2 = 1

Note. Bold text indicates a trial scored as inconsistent.

Table B2
Response consistency by experiment and consistency type

	Experiment 1	Experiment 2	Experiment 2B	Experiment 3
Visual strategy	2 chose R (16, 13) 1 chose L (13)	1 Chose R (15) 2 chose Melody 1's object (15, 15)	1 chose L (15) 2 chose Melody 1's object (13, 13)	–
Auditory strategy	1 reversed mapping (14) 1 locally consistent (12/15)	1 reversed mapping (13) 2 locally consistent (14/15, 14/15)	3 correct mapping (16. 13, 13) 2 reversed mapping (14, 13) 1 locally consistent (14/15)	18 correct mapping (16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 15, 15, 14, 14, 14)