The Development of Rhythmic Categories as Revealed Through an Iterative Production Task

Karli Nave[1,4*], Chantal Carrillo[2] , Nori Jacoby[3], Laurel Trainor[2], and Erin Hannon[4]

[1] University of Western Ontario, 1151 Richmond Street, Western Interdisciplinary Research Building, London, Ontario, Canada, N6A 3K7

[2] McMaster University, 1280 Main St W, Hamilton, ON, Canada, L8S 4L8

[3] Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

[4] University of Nevada Las Vegas, 4505 S Maryland Pkwy, Las Vegas, NV, United States, 89154

*Corresponding Author: Karli Nave, email: knave@uwo.ca, Address: Western University (ATTN: Karli Nave), 1151 Richmond Street, Western Interdisciplinary Research Building - Room 3190, London, Ontario, Canada, N6A 3K7

Declarations of interest: none

Acknowledgements: The first author was funded by the Top Tier Doctoral Research Fellowship from the University of Nevada Las Vegas during the completion of this work. We would also like to acknowledge the families who volunteered to participate in this study.

Highlights

- By age 6, children exhibit bias toward producing integer-ratio rhythm categories.
- Categorical rhythm priors closely resemble those of adults in the same culture.
- Age-related differences suggest less universal rhythm categories develop more slowly.
- Iterative tapping variability correlates with perception of rhythmic categories.
- Rhythm priors are likely acquired from passive experience prior to middle childhood.

Abstract

Both humans and non-humans (e.g. birds and primates) preferentially produce and perceive auditory rhythms with simple integer ratios. In addition, these preferences (biases) tend to reflect specific integer-ratio rhythms that are common to one's cultural listening experience. To better understand the developmental trajectory of these biases, we estimated children's rhythm biases across the entire rhythm production space of simple (e.g., ratios of 1, 2, and 3) three-interval rhythms. North American children aged 6-11 years completed an iterative rhythm production task, in which they attempted to tap in synchrony with repeating three-interval rhythms chosen randomly from the space. For each rhythm, the child's produced rhythm was presented back to them as the stimulus, and over the course of 5 such iterations we used their final reproductions to estimate their rhythmic biases or priors. Results suggest that regardless of the initial rhythm, after 5 iterations, children's tapping converged on rhythms with (nearly) simple integer ratios, indicating that, like adults, their rhythmic priors consist of rhythms with simple-integer ratios. Furthermore, the relative weights (or prominence of different rhythmic priors) observed in children were highly correlated with those of adults. However, we also observed some age-related changes, especially for the ratio types that vary most across cultures. In an additional rhythm perception task, children were better at detecting rhythmic disruptions to a culturally familiar rhythm (in 4/4 meter with 2:1:1 ratio pattern) than to a culturally unfamiliar rhythm (7/8 meter with 3:2:2 ratios), and performance in this task was correlated with tapping variability in the iterative task. Taken together, our findings provide evidence that children as young as 6-years-old exhibit simple integer-ratio categorical rhythm priors in their rhythm production that closely resemble those of adults in the same culture.

Keywords (max 6): rhythm, tapping, development, music cognition, iterated learning

1 Introduction

Music is characterized by temporal intervals having simple-integer ratios such as 1:1, 2:1, and 3:2, which play a fundamental role in structures such as key, harmony, rhythm and meter. When producing rhythmic patterns, both humans and some non-human species have shown a systematic bias towards discrete ratio categories, particularly 1:1 or isochrony, and this tendency towards discretization might be a key feature of culturally transmitted communication systems across species (De Gregorio et al., 2021; Ravignani et al., 2016; Raimondi et al., 2023; Roeske et al., 2020). Human musical systems nevertheless vary widely across cultures, and complex rhythmic ratios or even relatively simple-integer ratios such as 3:2 are more prevalent in some musical traditions than in others (Brown & Jordania, 2013; Polak et al., 2018; Savage et al., 2015). The observed biases towards discrete rhythmic categories presumably reflect a combination of learning constraints and exposure to specific patterns in an individual's musical environment during development.

The tendency for Western adults to impose simple-integer rhythmic categories onto an otherwise continuous range of possible ratios has been well documented. Simple-integer rhythmic ratios are prominent in music performance and spontaneous rhythmic tapping (Fraisse, 1982; Murton et al., 2017; Roeske et al., 2020), and listeners exhibit a strong tendency to assimilate more complex ratios into simple-integer ratios during rhythm reproduction (Collier & Wright, 1995; Cummins & Port, 1998; Essens & Povel, 1985; Jacoby & McDermott, 2017; Povel, 1981; Repp et al., 2012a, 2012b) and perceptual identification or transcription of rhythmic values to notation (Desain & Honing, 2003). Although Western music and listeners have been vastly overrepresented in the literature on rhythm perception and production, recent evidence provides strong support for the notion that listeners' biases towards specific ratios such as 3:2 do vary across cultures, and these variations map onto the relative prominence of those rhythmic ratios in participants' local musical traditions (Hannon, Soley, et al., 2012; Hannon & Trehub, 2005a; Jacoby et al. 2021; Jacoby & McDermott, 2017; Kalender et al., 2013; Polak et al., 2018; Ullal-Gupta et al., 2014).

Listeners presumably acquire culture-specific rhythmic categories during childhood, but much of the evidence for rhythmic categories in children comes from relatively indirect perceptual measures. For example, looking-time based habituation procedures suggest that at 12 months of age, North American infants can discriminate rhythmic variations of a song with a repeating 1:1:2 rhythmic pattern (which fits with 4/4 meter and is ubiquitous in Western music) but not variations of a song with a 3:2:2 rhythmic pattern (in 7/8 meter, which is found in Turkey, Bulgaria, and other regions but not common in Western music) (Hannon & Trehub, 2005b). By contrast, 6-month-old infants can discriminate variations of either type of song (Hannon & Trehub, 2005a), but they fail to discriminate same size disruptions to a song with a more complex 7:4:4 rhythm (Hannon et al., 2011). Robust listening preferences for culturally familiar rhythmic structures have also been observed in 4- to 8-month-old infants, with North American infants preferring 2:1:1, compared to 3:2:2 and 7:4:4, and Turkish infants preferring either 2:1:1 or 3:2:2, compared to 7:4:4 rhythms (Soley & Hannon, 2010).

Together, the above infant findings could be taken as evidence that adult-like rhythmic categories are acquired by one year after birth. However, other evidence suggests that rhythm perception and production continue to change in important ways throughout childhood. In perceptual tasks, young North American children exhibit an advantage for rhythmic structures with 2:1 versus 3:2 ratios (Einarson & Trainor, 2016; Hannon, Der Nederlanden, et al., 2012) However, brief at-home exposure to music with 3:2 ratios reduces or eliminates these biases among young children but not older children or adults, implying weaker or less stable rhythmic categories in younger listeners (Hannon, Der Nederlanden, et al., 2012; Hannon & Trehub, 2005b). Other perceptual paradigms suggest that the ability to match a metronome to music or to sustain an internal sense of the beat does not approach adult-like levels until early adolescence (Nave et al., submitted; Nave-Blodgett et al., 2021b, 2021a). This gradual developmental trend is consistent with findings from production paradigms, where adult-like levels of synchronization to music and other rhythmic stimuli do not emerge until late childhood or early adolescence (McAuley et al., 2006; Monier & Droit-Volet, 2019; A. T. Tierney & Kraus, 2013).

It is perhaps because children have such difficulty with production tasks that much of the developmental research on rhythmic categories comes from perceptual paradigms that measure rhythm categories indirectly. However, this limits direct comparison of children with adults because even when doing the same task, adults and children may succeed or fail for different reasons. For example, just because adults tend to use the perceived beat as a reference point for a rhythm discrimination task does not necessarily mean that infants or children also do this (Gordon et al., 2015; Hannon et al., 2018; Ozernov-Palchik et al., 2018). Another limitation of discrimination paradigms is that they typically can only test two or three rhythmic structures at a given time, which requires the researcher to impose assumptions about which rhythm categories are most important and worth measuring. A production paradigm that presents listeners with a wide range of rhythmic patterns and yet is simple enough to be performed by children allows for more direct measurement and a more thorough understanding of the development of rhythm categories.

It is for this reason that we adapted a paradigm employed previously (Jacoby & McDermott, 2017), which uses iterated tapping to reveal listeners' *rhythm priors*, or internal expectations about rhythmic categories. In this task, listeners were first presented with a random 3-interval "seed" rhythm that they attempted to synchronously reproduce. Their reproduction was then presented back to them as the next target stimulus (similar to the game of "telephone") and, over multiple iterations, responses converged on integer-ratio rhythms (Jacoby & McDermott, 2017). Crucially, the integer-ratio rhythms produced after five such iterations of a seed rhythm were dependent on culture, with U.S. adults more likely to converge on culturally-familiar ratios such as 1:1:2 and 3:3:2, whereas listeners from a native Amazonian society converged less on some ratios such as 3:3:2 but still showed strong 1:1:1 and 1:1:2 categories. More recently, a large-scale cross-cultural study of 39 participant groups in 15 countries used the iterated tapping paradigm to provide robust evidence for global variation in rhythmic categories such as 3:3:2 and 3:2:2 that could be mapped to local musical traditions for different groups (Jacoby et al., 2021). Importantly, because not all integer ratios were favored by all participants from all cultures, this provides support for the important role of enculturation. Differences between cultures in adult responses also raise the question of when culturally specific rhythmic categories develop and if some categories develop earlier than others. The fact that adults from different cultures varied so much implies there may be significant differences between children and adults, even within the same culture.

Because the iterated tapping paradigm is fairly intuitive and relies minimally on verbal instruction, it has been successfully used with a diverse range of participants with different education levels and familiarity with experimental testing methods and environments. We therefore expected that children would also be able to perform the task. Because the paradigm design optimizes group characteristics over individual participant characteristics, it allows for the entire rhythm production space to be measured and biases to be estimated separately for different age groups. We expected that children, like adults, would show a bias for producing integer-ratio rhythms such as 2:1:1 as shown in prior work (Jacoby et al., 2021; Jacoby & McDermott, 2017). If rhythmic categories are acquired early in development, as suggested by perceptual paradigms, we expected children's rhythmic categories for other ratios such as 3:3:2 and 2:2:3 to be similar to those demonstrated by U.S. adult participants. However, we also predicted that the youngest children might show less robust categories that would become stronger with age. For comparison with prior work, we also included a perceptual measure that required children to make similarity judgments about rhythmic variations of songs with culturally familiar and culturally unfamiliar rhythms, as used in prior work (Hannon, Der Nederlanden, et al., 2012; Hannon & Trehub, 2005a). This allowed us to examine whether a greater advantage for culturally-familiar rhythms in the perceptual task would be related to better motor synchronization with culturally-familiar integer ratio rhythms in the iterated tapping task.

2 Methods

2.1 Participants

A total of 158 children were recruited from the Las Vegas, Nevada and Hamilton, Ontario communities. Children were recruited in three age groups. Fifty-six children aged 6 to 7 years (M = 6.89, SD = 0.52), fifty children aged 8 to 9 years (M = 8.88, SD = 0.62), and fifty-two children aged 10 to 11 years (M = 11.00, SD = 1.81) completed the experiment. All participants provided assent and their parents provided permission in accordance with the University of Nevada, Las Vegas Institutional Review Board or the McMaster Research Ethics Board. Musical experience was estimated for each child using the maximum number of parent-reported formal years of training for either music (instrument or voice) or dance. See Table 1 for sample sizes broken down by age group, testing location, and task completed, and for demographic information by age group. The sample size (number of overall iterations) was established based on an earlier publication (Jacoby & McDermott, 2017). Note that all child age groups did not differ from one another on the basis of estimated IQ (WASI scores) or socio-economic status (estimated by parent education). See Table 1 for more information on demographics.

2.2 Experimental Design

The experiment consisted of four parts: 1) the Iterative Tapping (IT) task, 2) the Rhythm Perception (RP) task, 3) Tapping Baseline (TB) Task, and 4) the Vocabulary and Matrix Reasoning of the Weschler's Abbreviated Scale of Intelligence (WASI)[1]. Each of these components is described in detail below. Participants at McMaster University completed only tasks 1 and 3 and could therefore finalize the experiment in one session. Participants at UNLV completed all four tasks across two sessions, with some attrition (25 children of 91 overall) where children did not return for visit 2 and thus are missing data for some tasks. Total number of participants included for each task is reported in Table 1.

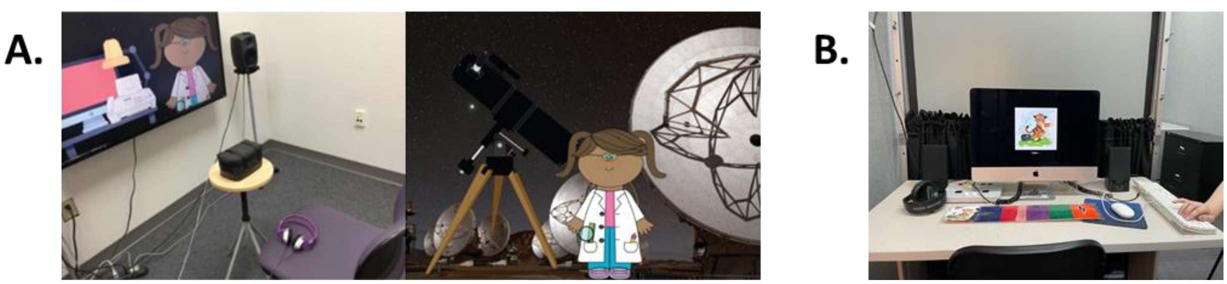

Figure 1. Experimental paradigm. (A) Iterative Tapping "Alien" Game. Left: lab setup. The participant sits in front of a large screen. A tapping response box (black box positioned on the table) is used to collect responses. Right: children's game design. (B) Rhythm perception "Animal" Game. The participant sits in front of a computer screen and makes responses by moving their game piece along a colored-square board. The experimenter enters the participant's responses on the keyboard.

[1]Note: Only a subset of the participants from the Las Vegas area completed tasks 2 and 4. See Descriptives in Table 1.

2.3 Iterative Tapping (IT) Task

The IT task was adapted from Jacoby and McDermott (2017). We created a child-friendly computer game, in which the child acted as a junior astronomer assisting Estelle, a scientist attempting to communicate with aliens, by tapping in synchrony with repeating three-interval rhythms sent to Earth from outer-space (see Figure 1). The child was instructed to begin tapping with each rhythm as soon as possible and continue tapping until the rhythm stopped playing. On each trial, the child tapped to a random "seed" target rhythm for the first iteration, and on each of 4 subsequent iterations the child tapped to the rhythm they just produced on the previous iteration (see stimulus details below).

After each trial, the aliens would "respond" to the communication by sending another message, which served as a noncontingent reward. Over the course of the 10 trials, each child received a "we come in peace" message, 4 pieces to spaceship assembly instructions, 4 pieces to a map to the alien's planet, and a dictionary to learn to speak the alien's language. At the conclusion of the game, the children "travel" through space to visit the aliens with Estelle. The experiment script (MATLAB) and the accompanying computer game (Powerpoint) can be found on our OSF page: https://osf.io/6ugrb/.

	Age 1: 6-7 Years (N=56)	Age 2: 8-9 Years (N=50)	Age 3: 10-11 Years (N=52)
UNLV: McMaster[1]	30:26	30:20	47:5
TB sample	n=43	n=47	n=47
RP sample	n=28	n=25	n=42
WASI sample	n=26	n=24	n=34
IT sample	n=51	n=43	n=44

IT num. blocks (excluded)	473 (37)	404 (26)	422 (18)
Number of raw taps (excluded)	46910 (13029)	40495 (7434)	39853 (5846)
Females: Males	27:29	27:23	23:29
Mean Age	6.9±0.52	8.9±0.78	11.0±0.71
Music Experience[2]	1.04±1.41	1.52±2.08	1.58±2.06
SES[3]	4.87±1.24	4.70±1.44	4.98±1.13
WASI Composite Score	99.96±16.11	109.83±15.35	102.24±17.78

Table 1 *Participant Sample Sizes & Demographics.*

[1]Due to the COVID-19 pandemic, data collection was halted indefinitely for McMaster University. More data was collected at UNLV because local health regulations allowed them to open sooner.

[2]Music Experience scores were calculated as the maximum number of years of formal training in either music (instrument or voice) or dance.

[3]SES Scored on a 6-point scale based on highest level of parental education (1: No high school, 2: High school diploma, 3: Some college, 4: Technical School, 5: Four-year college/university degree, 6: Graduate degree).

Table 1. Descriptives for participants by age group. Top: sample size broken down by age group, then further by testing location and task. TB= Tapping Baseline task. RP = Rhythm Perception task. WASI=Weschler's Abbreviate Scale of Intelligence. IT= Iterated Tapping task. IT num. blocks = Number of usable blocks (1 block = 1 seed rhythm with 5 tapped iterations), excluded blocks indicated in parentheses. Note that not all participants completed all tasks, as some attrition occurred before session 2 for some participants at UNLV. Bottom: descriptive data for each age group.

2.3.1 Apparatus.

All stimuli were presented through Sennheiser HD 280 Pro headphones (UNLV) or Creative Inspire T10 2.0 Speaker System (McMaster) at a comfortable listening level (approximately 60 dB). The experimenters constructed a tapping sensor to record finger tapping responses, which was made from soft material (foam) to provide minimal auditory feedback (Figure 1A). Inside the apparatus, a microphone was attached to the surface of the sensor, allowing sensitivity to the lightest touch by the participant. Data were acquired with a Focusrite Scarlett 2i2 USB sound card, which simultaneously recorded the microphone tapping input and the headphone rhythmic sound stimulus output. Stimulus and response (tapping) onsets were extracted from the stereo audio signal using a MATLAB script. This apparatus was used in previous studies (Jacoby and McDermott 2017, Jacoby et al. 2021, Anglada-Tort et al. 2022) and provides high tapping precision (<5ms jitter and latency; see Anglada-Tort et al. 2022).

2.3.2 Stimuli.

The rhythmic patterns were comprised of short percussive sounds ("clicks") lasting 55 ms with an attack time of 5 ms. For each iteration in a trial, we presented a click stream originating from a three-interval rhythm (s_1, s_2, s_3), which was repeated ten times continuously. When the rhythm is cycled, the fourth event forms the end of the last interval and the beginning of the first interval (Figure 2C). The overall duration of each three-interval rhythm was fixed to 2000ms, making a one-to-one mapping between all three-interval rhythms possible. Because the overall duration is fixed, the intervals can be presented on a two-dimensional triangular map (Figure 2B) known as a ''chronotopological map'' (Desain & Honing, 2003; Honing, 2013; Jacoby & McDermott, 2017) – for simplicity, here we refer to it as the ''rhythm triangle.'' Every three-interval rhythm with a fixed duration can be uniquely described by a point on the triangle.

The coordinates of the rhythm in 2D space are a linear combination of the vertices of the triangle (P_1,P_2, P_3 in Figure 2A), so for example a rhythm of 1:1:2 with intervals that are ¼ , ¼ , ½ of the cycle duration is located at location $\frac{1}{4} P_1 + \frac{1}{4} P_2 + \frac{1}{2} P_3$. Points that are very close to the vertices represent rhythms with at least one very short (near 0 ms) interval. In order to avoid presenting rhythms with intervals too short to reproduce (London, 2012), we constrained intervals in the seeds to a smaller simplex (inner triangle in Figure 2B) such that all intervals were longer than 300 ms (15% of the total duration). Thus, all intervals in the initial seeds were effectively constrained not to exceed the perception / production limit (about 100–200 ms in adults, Repp, 2005) (see Figure 2B).

Each initial seed rhythm consisted of a randomly selected point on the inner triangle (uniformly distributed), such that the rhythm presented corresponded to the three-interval rhythm (s_1, s_2, s_3). Three-interval rhythms were formed by repeating a sequence of three time intervals defined by four events (e.g., clicks). The seed rhythm repeated for ten repetitions, and after a few clicks (typically a bit more than one cycle), participants began to synchronize to the click stream. Using the participant's inter-response intervals on the first iteration, we obtained an averaged three-interval pattern ($\underline{r}_1$, $\underline{r}_2$, $\underline{r}_3$) to present as the target rhythm of the next iteration, and this process was repeated for five iterations per trial (see Figure 2C).

We excluded trials with too many missing tap onsets. Each tapping onset was associated with the most proximal stimulus onset. We allowed some missing tapping onsets within each repeated three-interval pattern. However, we required that each of the three stimulus onsets within the seed be associated with a tap onset in at least three of the ten repetitions, and that the average response ($\underline{r}_1$, $\underline{r}_2$, $\underline{r}_3$) was not located far beyond the inner simplex, i.e., did not contain an interval shorter than $0.95f$ of the overall duration (where f determines the boundary of the inner simplex as described above; here this constraint excludes intervals shorter than 285 ms). If the iteration satisfied the two criteria above, we set the new seed pattern to the response pattern: (s_1, s_2, s_3) $\leftarrow$ ($\underline{r}_1$, $\underline{r}_2$, $\underline{r}_3$). Additional iterations were not added to replace invalid iterations; if an iteration was invalid, the seed remained unchanged, and the data from that iteration was omitted from analysis. Participants each completed 10 trials (where each trial is defined as a set of 5 iterations derived from a single seed). Details regarding excluded and included trials by age group can be found in Table 1. For demographics broken down by each lab, see Supplementary Table 1.

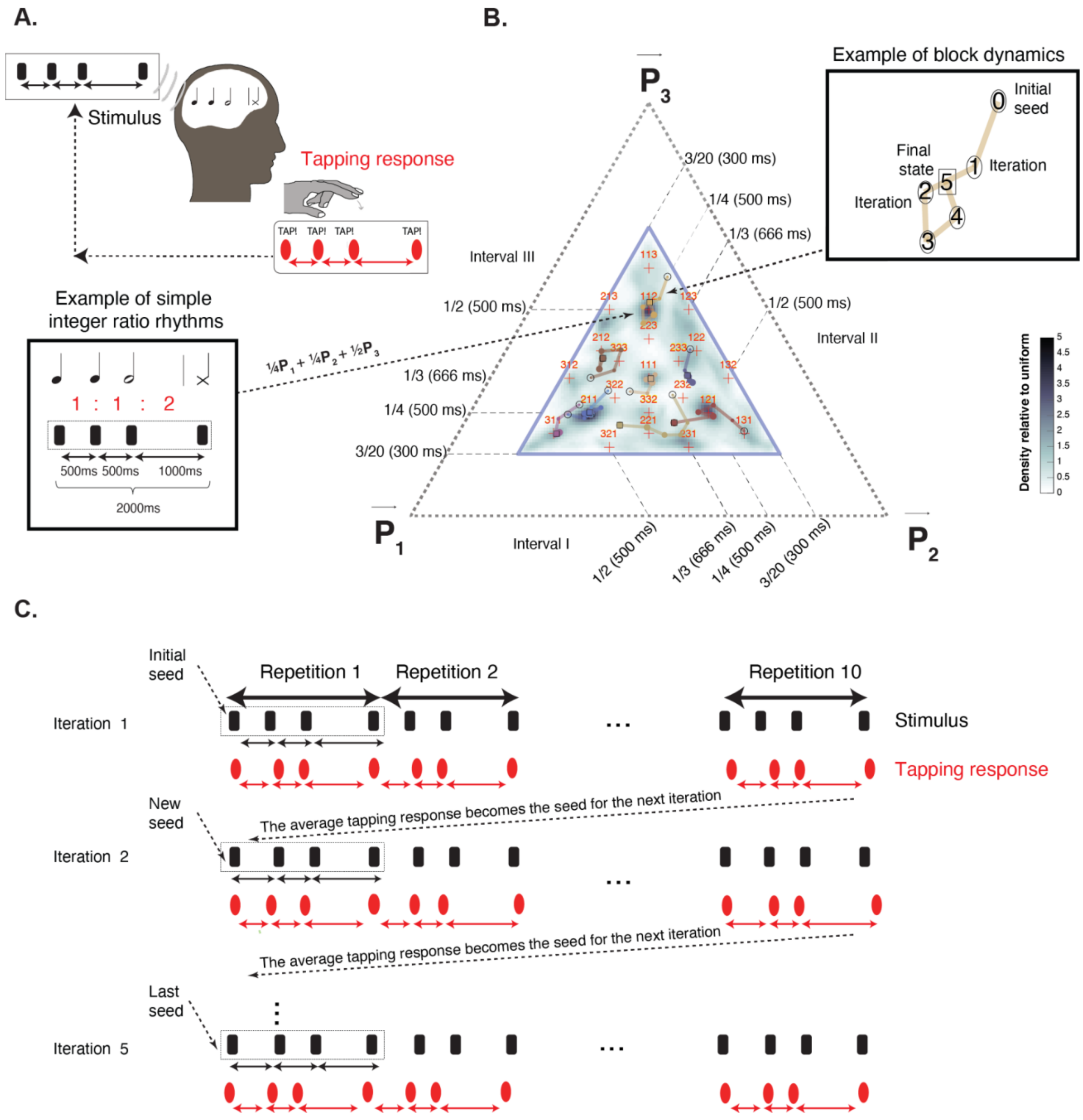

Figure 2. Illustration of stimuli and trial structure for the iterative tapping (IT) task. (A) Schematics of the experimental paradigm. On each trial, the participant is presented with an initially random seed rhythm and they reproduce the rhythm by tapping along. Their reproduction becomes the stimulus on the following trial. This is repeated for 5 iterations. Each

participant followed this procedure for 10 initial random seed rhythms (i.e., completed 10 trials). (B) Triangular space representations. Every three-interval rhythm with a fixed duration can be uniquely described by a point on the triangle, that is the linear combinations of the vertices of the triangle (P_1,P_2, P_3). The inner blue triangle represents the range of rhythms considered for the experiments (constrained so that all intervals are longer than 300 ms). The left inset shows an example of a simple integer ratio rhythm (1:1:2). The right inset shows the progression of an example trial. (C) Schematic of trial structure. The dark squares indicate onsets of the stimulus rhythm. The light-colored circles indicate the participants' attempted synchronized tapping responses. On each subsequent iteration, the average of the participants' responses on the prior iteration becomes the new seed stimulus rhythm. This continues for 5 iterations, which together form 1 trial for the experiment.

2.3.4 Onset Extraction.

In order to extract tapping onsets, we performed an extraction pipeline that is similar to previous studies (Jacoby and McDermott 2017; Anglada-Tortt et al. 2022). First we scanned the audio file in 15 second windows, detecting the earliest samples exceeding a relative threshold of 1.45% of the maximal power of the recorded waveform in the window and separated from the previous candidate by more than 80 ms. This procedure was applied to allow for some variation in the overall tapping amplitude within the trial, such as if the participant tapped louder at the beginning of the trial compared with later on. Second, we discarded responses that were too far from any stimulus click, regarding them as errors. Here we took into account an important characteristic of human tapping known as "negative mean asynchrony" (Repp 2005), namely that tapping in time to a beat tends to be at a slightly different phase compared to the beat onset (typically occurring before the beat). We first matched each tapping onset to its closest stimulus click and computed the mean asynchrony m as the average difference between a response and its corresponding stimulus click. If there were two events corresponding to the same target onset we choose the closest only. We then excluded all events such that $|(R - m) - S_i| > 150$ ms (namely, a window of 300 ms centered around the perceptual center defined by the mean asynchrony) from further analysis. The number of raw and excluded taps for each group are detailed in Table 1.

2.3.5 Data Analysis and Statistics.

One assumption of the iterated paradigm is that tapping behavior should stabilize, or converge, across iterations as the averaged production moves closer to the producer's internal rhythm prior. To first examine whether children reached convergence by the fifth iteration, we examined *copying error*. *Copying error* was calculated as the distance between stimulus and reproduction on each iteration, and it is expected to decrease across iterations as the participant's tapping stabilizes, typically by converging on rhythm priors (Jacoby & McDermott, 2017; Jacoby et al. 2021). To further understand the temporal dynamics underlying children's tapping behavior in the iterated task, we also calculated *tapping variability* as the standard deviation of the average distance between taps and the stimulus. Two repeated measures ANOVAs were used to estimate the effect(s) of Iteration and Age Group on copying error and tapping variability. Post hoc pairwise comparisons were conducted as necessary.

To estimate the presence of rhythm priors, we employed the same kernel density estimation procedure utilized by Jacoby and McDermott (2017). Knowing each rhythm was repeated J = 10 times, we assumed that the response on the jth repetition $(r_{i,m}^{j,1}, r_{i,m}^{j,2}, r_{i,m}^{j,3})$ was taken from a Gaussian distribution: $(r_{i,m}^{j,1}, r_{i,m}^{j,2}, r_{i,m}^{j,3}) \sim N\,(\mu_{i,m}, C_{i,m})$. Here $r_{i,m}^{j,k}$ is the inter-response interval for click k of repetition j of iteration m of trial i. We estimated the mean and covariance of the Gaussian for each iteration and trial and then summed all Gaussians to yield an estimate of the distribution. Since the covariance of the responses are estimated from a small number of samples, we regularized the empirical covariance matrix by adding to the empirical covariance matrix (Σ) a diagonal matrix ($\Sigma' = \Sigma + \sigma^2 I$, where σ=15 ms and I is the identity matrix). We used the regularized matrix Σ' as an approximation of the prior in the local neighborhood of the stimulus. This procedure is equivalent to using only the means of all responses from all iterations and performing a standard kernel density estimate (KDE) with the kernel width computed from the local estimates. To obtain the full kernel density estimate, we aggregated the Gaussians for each of the five iterations across all trials. We performed this same KDE estimated for each of the age groups: 6-7 years, 8-9 years, 10-11 years, and Adults[2].

To quantify the similarity of the resulting KDE distributions across groups, we calculated Jensen-Shannon divergence (JSD) (Jacoby & McDermott, 2017; Jacoby et al., 2021; Majtey et al., 2005; Wong & You, 1985). Since JSD is always a positive number (between 0 and 1), it is expected to be different from 0 when the kernel density estimates being compared are determined based on a finite sample. We used bootstrapping to estimate whether the distance between groups, as estimated by JSD, is greater than what would be expected based on this finite-sampling effect. To do this, we created 1000 simulations where we split each group into two randomly-sampled halves (50% of trials in each half) and performed 1000 kernel density estimates on each half. We then computed JSD for each split half, comparing JSD within-groups (comparing split halves from the same group) to the JSD across-groups (comparing split halves across comparison groups). Namely, to declare that two groups are significantly different, their mean JSD difference had to be significantly greater across-groups than within-groups. To further explore differences among age groups, we examined whether the vector of category weights produced by the KDE analysis were related across groups using Pearson's correlations.

[2] Adult data (n=26) was shared from Jacoby & McDermott (2017) Experiment 1, and was used with permission for comparison purposes. It includes n=13 musicians (9-23 years experience) and n= 13 non-musicians (0-2 years experience) from the Boston area.

To see if the results are significantly close to integer ratios, we computed an integer score. First we defined the set of simple integer ratios in Figure 2 to be all three-interval integer ratios involving only the numbers 1,2, and 3. This results in 21 categories: 1:1:1, 1:1:2, 1:2:1, 2:1:1, 1:2:2, 2:1:2, 2:2:1, 1:1:3, 1:3:1, 3:1:1, 1:2:3, 2:3:1, 3:1:2, 1:3:2, 2:1:3, 3:2:1,2:2:3, 2:3:2, 3:2:2, 2:3:3, 3:2:3 and 3:3:2. To compute the overlap, we calculated an "integerness score" by computing the average minimal distance between responses in the final iterations and the set of 21 points. We compare this to a distribution of distances obtained where nth set of 21 points were selected at random (uniformly from the triangle).

To measure the relative weight (importance) of each category within each group, we used a Gaussian mixture model in which the mean of each mixture component was constrained to be close to a small-integer-ratio rhythm. To guarantee that each mode was associated with the same category across groups, we required that the category center be constrained to a distance that is less than 1/2 of the minimal distance between the 21 integer categories mentioned above. This analysis yields estimates of the *category weights*, or the relative density of the distribution underlying each rhythm category, such that each of the 8 integer-ratio categories (collapsed across cyclic permutation of the same rhythmic pattern; e.g., we computed together the average weights of 1:1:2, 1:2:1, and 2:1:1) received a weight between 0 and 1, where the collective category weights across the entire distribution sum to 1. Category weights were compared across the four different age groups (6-7 years, 8-9 years, 10-11 years, Adults) using independent samples t-tests Bonferroni-corrected for multiple comparisons.

2.4 Rhythm Perception (RP) Task

We also administered the Rhythm Perception (RP) task from Hannon et al. (2012), which required children to listen to a standard song and rate the similarity of subsequent variations on that song. The task was presented to children as a game, in which the child was asked to judge how well various animals matched an original song performed by a tiger musician. Participants made their judgments by placing a game piece on a colorful horizontal board with 6 different squares, depicting the tiger on the left-most square (see Figure 1B). Children were instructed to put their game piece directly on the tiger if the animal reproduced the song perfectly, close to the tiger if it almost matched, and far away from the tiger if the animal did not at all match the Tiger's song.

To ensure children understood the task, each child first completed a block of six training trials with a familiar tune, "Mary had a little lamb." They were provided feedback after each response, indicating the ideal/expected rating for each practice song ("most people would say that song was similar to the tiger's song and place the game piece next to the tiger"). Children then completed four test blocks, two of which presented songs with culturally familiar rhythms (in 4/4 meter) and two of which presented songs with culturally unfamiliar rhythms (in 7/8 meter). Songs in 4/4 meter contained rhythmic patterns with a 2:1:1 or 1:1:2 structure, and songs in 7/8 meter contained rhythmic patterns with a 2:2:3 structure (see Figure S1). The order of song blocks was counterbalanced between subjects and the order of variations on each trial was determined randomly. The entire task took approximately 10-15 minutes.

2.4.1 Apparatus

All stimuli were presented to the participant through Sennheiser HD 280 Pro headphones at a comfortable listening level (approximately 60 dB). The experiment was designed and presented using Presentation software (Neurobehavioral Systems, Inc., 2016). Responses were recorded by the experimenter using the keyboard numbers 1-6, where 1 was on the Tiger square, 2 was 1 square to the right of the tiger square, etc.

2.4.2 Stimuli

The auditory stimuli were identical to those used in Hannon et al. (2012). All stimuli were generated using a MIDI sequencer (Digital Performer) and converted to AIFF using the Apple QuickTime synthesizer (Apple, Inc., Cupertino, CA). Each trial consisted of an 8-measure standard “tiger” song, followed by 4 “animal” variations of that standard. Standard songs consisted of four traditional Eastern European folk songs, two in 4/4 isochronous meter, and two in 7/8 non-isochronous meter. Each standard stimulus had two melodic instruments playing in unison or thirds, a harmonic instrument, and a fourth percussion instrument. The percussion instrument consisted of a long-short-short or a short-short-long rhythmic pattern that was 2:1:1 or 1:1:2 for the 4/4 songs and 3:2:2 or 2:2:3 for the 7/8 songs. Because the eighth-note duration was set to 250 ms, each cycle of the 3-interval percussion rhythm was 2000 ms for 4/4 meter songs and 1750 ms for 7/8 meter songs.

For each standard there were four variations, simplified to only one melodic instrument (a piano playing the main melodic line) and one percussion instrument (a wood block playing the long-short-short or short-short-long drum patterns). Two variations preserved the original meter and drum pattern of the standard stimulus. One “unaltered” variation presented exactly the same duration notes in both the piano and drum part, while the other “structure-preserving” variation added an extra note to the piano melody but reduced the duration of the neighboring note so as not to disrupt the meter or drum pattern. The other two variations disrupted the meter and drum pattern. The “structure-disrupting” variation added a full 250-ms note to the long duration in each measure, thus increasing each measure or cycle of the drum pattern by 250-ms and resulting in a 2:1:1 drum pattern for the 7/8 songs and 2.5:1:1 for the 4/4 songs. The "severely disrupted" variations had pseudorandom insertions of two extra notes or durations per measure and resulted in a highly variable drum pattern. All stimuli were accompanied by a cartoon display, with the standard stimulus accompanied by a tiger holding a guitar and each variation accompanied by a different animal holding an instrument. Stimulus wav files can be found on the project OSF page: https://osf.io/6ugrb/.

2.4.3 Data Analysis

Accurate performance was reflected by giving higher similarity ratings to structure-preserving than to structure-disrupting test stimuli. In line with Hannon et al. (2012), we calculated children's *perception score* by subtracting their ratings of the test variations that preserved the meter (the mean of unaltered and structure-preserving ratings) from ratings of test stimuli that disrupted the meter (mean of structure-disrupting and severely disrupted ratings). A positive score indicates more accurate performance. For each child we calculated an average perception score for 4/4 meter and for 7/8 meter.

2.5 Tapping Baseline (TB) Task

To obtain basic motor synchronization measures, each participant completed the Tapping Baseline (TB) task, which comprised three trials. On the first trial, we measured spontaneous motor tempo by asking participants to tap during silence at a comfortable speed that was "not too fast and not too slow." On the second trial, participants synchronized their taps with a slow 800 ms metronome, and on the third trial they synchronized with a faster 600 ms metronome. All stimuli were presented and recorded in Audacity ® 2.4.2 (Audacity Team, 1999-2021). Individual .wav files were extracted from Audacity for each trial and then analyzed for tap onsets using MATLAB. Spontaneous motor tempo was calculated by taking the median tempo for Trial

1. Asynchrony scores were calculated by taking the relative average deviation of taps from the stimulus, averaged across Trial 2 and Trial 3. To estimate variability of tapping, we calculated the coefficient of variation (standard deviation of produced intervals/target interval), then averaged for Trials 2 and 3. A univariate ANOVA was used to estimate the effect of age group on Spontaneous Motor Tempo, Tapping Baseline variability, and Tapping Baseline Asynchrony.

2.6 WASI

Finally, the Vocabulary and Matrix Reasoning subtests of the WASI were administered. In the Vocabulary subtest, children are given a word verbally and asked to provide a definition. In the Matrix Reasoning subtest, children view an incomplete matrix or series of pictures and are asked to select the option that completes the matrix. Taken together, these two subtests are used to determine the Full Scale 2 composite score, which is age-normed and serves as an estimate of the child's IQ. We estimated age-normed IQ scores to ensure no significant differences among our three age groups. Mean WASI Composite Scores and standard deviations are provided in Table 1.

2.7 Relation Between Perception and Production

To our knowledge, this is the first time an iterative tapping procedure has been used to estimate rhythm priors in children. For this reason, we were interested in validating the measurement against more basic measures of tapping behavior in children. Tapping variability has been used as a proxy for motor entrainment across the lifespan (McAuley et al., 2006; Thompson et al., 2015), and it offers significant predictive power for other cognitive abilities such as reading (Carr et al., 2014; Dellatolas et al., 2009; A. T. Tierney & Kraus, 2013), sustained attention (A. Tierney & Kraus, 2013), and neural speech encoding (Carr et al., 2014). Furthermore, because each child received different seed rhythms selected at random, and because the space of possible seed rhythms was large, some children may have received more difficult rhythms than others, making estimation of individual differences from IT potentially problematic. We therefore examined whether tapping variability in the IT task was related to the more traditional tapping variability measure from the TB task by conducting correlations among the IT and TB tasks after controlling for age.

In addition (assuming variability was correlated across IT and TB tasks), we were interested in investigating whether individuals' tapping variability on the IT task was related to their performance on the RP task. We hypothesized a relation between RP scores for 4/4 meter with tapping variability on the IT task for rhythm categories mapping onto 4/4 meter – namely, 2:1:1 and 3:3:2. Similarly, we hypothesized a relation between RP scores for 7/8 meter with tapping variability on the IT task for rhythm categories mapping onto 7/8 meter, which is 2:2:3. To this end, we examined partial correlations controlling for age between *production* (IT tapping variability) and *perception* (RP scores).

2.8 Exploratory Analyses

In addition to our main analyses, we conducted exploratory KDE analysis and group comparisons based on two other grouping variables: music/dance experience and region/ethnic background. The exploratory analyses did not include the Adult participants. To estimate the effect of music and dance experience on underlying rhythm priors, KDE estimates were initially performed on three groups of children: those with 0 years, 1-2 years, or 3 or more years of musical experience. Group comparisons were conducted with post-hoc examinations of JSD and by comparing category weights.

To estimate the effect of region/ethnicity, KDE estimates were initially performed on three groups of children: self-reported Hispanic children from UNLV, age-matched Non-

Hispanic children from UNLV, and age-matched Non-Hispanic children from McMaster. We focused on this comparison because one of the largest differences among the two testing locations was the ethnic background of our samples. According to recent census data, 31% of people who live in Las Vegas are Hispanic, while less than 2% of people who live in Hamilton, Ontario are Hispanic. This provides an interesting sample of convenience because individuals who self-identify as Hispanic might have more exposure to Latin, Afro-Cuban, and Afro-diasporic music traditions where the “tresillo” and clave rhythms prominently feature a 3:3:2 pattern. Because we also expected all participants in Las Vegas, NV (regardless of language background) to have relatively greater exposure to Latin music than the McMaster group, we also included region as part of the grouping criteria. Group comparisons were conducted with post-hoc examinations of JSD and by comparing category weights.

3 Results & Discussion

3.1 Iterative Tapping (IT) Task

3.1.1 Aggregated Responses For all Participants. Figure 3 shows aggregated responses from all child participants who completed the IT task, collapsed across age group (n = 139; average of 7-8 seeds per individual, 1299 total distinct seeds), shown as points on the rhythm simplex for each iteration. The distribution of responses evolves over iterations and appears to converge to a multi-modal distribution. Specifically, the modes are hypothesized to correspond to integer ratio rhythms, which would indicate the presence of underlying rhythm priors (Jacoby & McDermott, 2017). We henceforth refer to these modes as “categories” because they demonstrate characteristics of categorical perception, biasing the perception of nearby rhythms toward the mode center (Feldman et al., 2009; Jacoby & McDermott, 2017).

Copying error was examined first to estimate whether or not, like adults, children's tapping stabilized over the course of the five iterations. We submitted the dependent measure of copying error to a 5 (Iterations) x 3 (Age Group) mixed design ANOVA, which yielded a significant Iteration x Age Group interaction, $F(8,540) = 3.068$, $p = .002$). Post hoc tests revealed that for all age groups, there was a significant decrease in copying error from iterations 1 to 2 and from iterations 2 to 3 ($p < .05$), and for ages 6-7 years and 10-11 years there was a significant decrease in copying error from iterations 3 to 4 ($p < .05$). For all age groups, copying error did not significantly differ between iterations 4 and 5 ($p > .05$), suggesting convergence was achieved by all age groups (Figure 4).

A 5 (Iterations) x 3 (Age Group) mixed design ANOVA on tapping variability yielded a main effect of Age Group ($F(2,135)=8.809$, $p < .001$) and a main effect of Iteration ($F(4,132)=7.413$, $p < .001$) with no interaction. Post hoc tests revealed a significant decrease in variability from iteration 4 to iteration 5 ($p < .05$), with no other significant differences among consecutive iterations. Independent t-tests revealed that 6- to 7- year-olds had significantly higher tapping variability than 8- to 9- year olds ($p < .05$) and 10- to 11- year-olds ($p < .05$), but no difference between 8- to 9- year-olds and 10- to 11- year-olds ($p > .05$). This is consistent with prior work showing that children's ability to synchronize their movement with rhythm continues to improve with age (McAuley et al., 2006; Monier & Droit-Volet, 2019).

3.1.2 Kernel Density Estimation of Rhythm Priors. The responses were concentrated in a relatively small area. Quantitatively, 33% of the area contained more than 50% of the density. Moreover, modes of KDE increased density in the distributions occur on simplex locations corresponding to integer ratios, as indicated by the red plus symbols in Figure 5A. Quantitatively we can compute an "integerness score" - a measure of closeness of the responses to integer points - and compare this score with the baseline, representing closeness to random locations. In all cases, we found highly significant integerness scores (see Figure 4C).

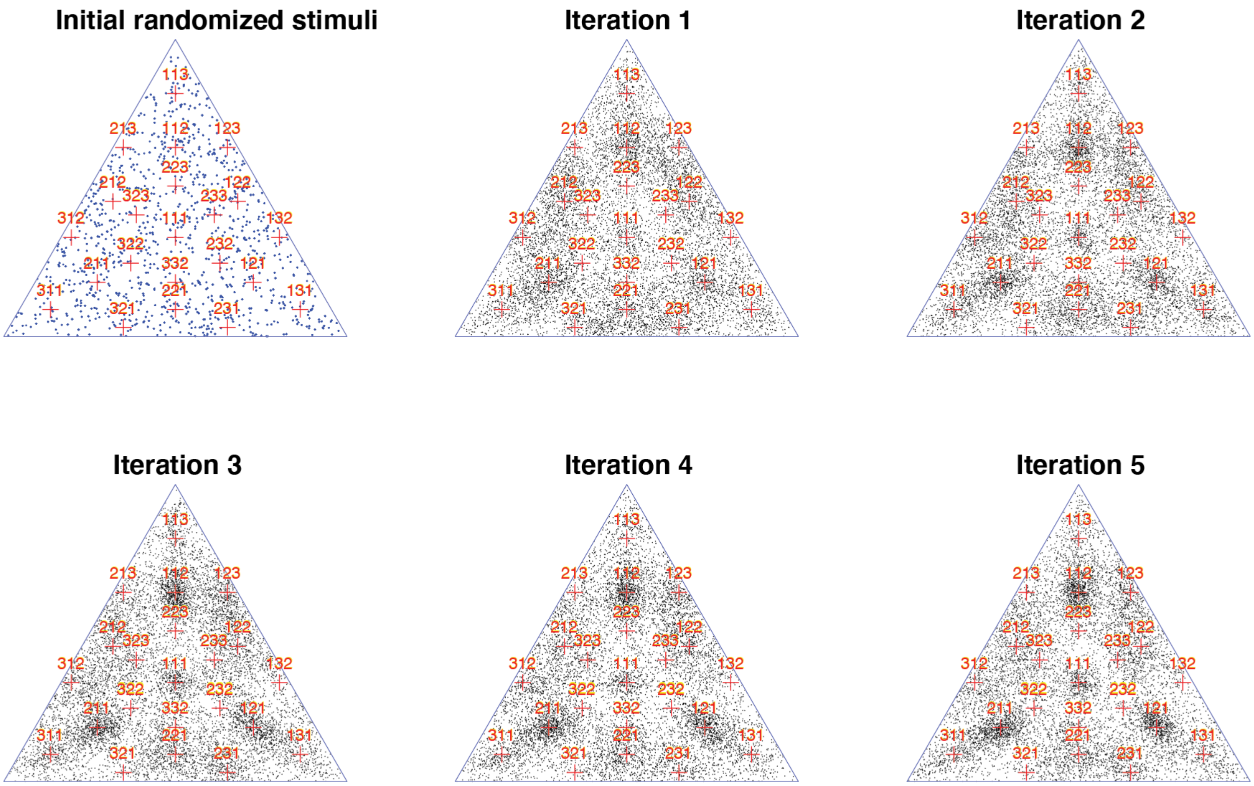

Figure 3. Initial randomized stimulus (seed rhythm), followed by the iterated reproduction by synchronous tapping across all iterations, collapsed across age groups. Each point on the simplex represents the averaged response taken from the participants' synchronized tapping.

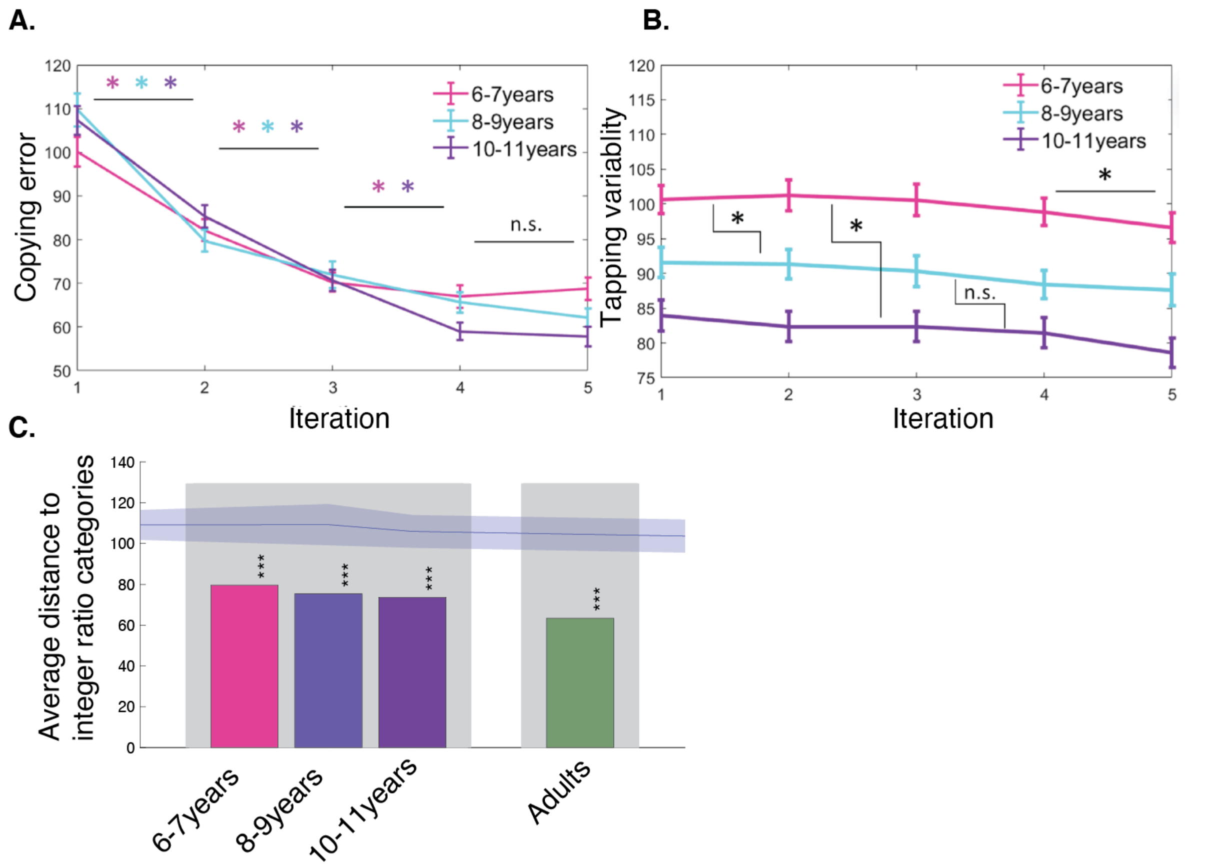

Figure 4. Results from the iterative tapping (IT) task. (A) Copying error (in milliseconds) by age group and iteration. * = indicate a significant ($p < .05$) post-hoc pairwise comparison of copying error between successive iterations for each age group. (B) Tapping variability by age group and iteration. * = significant ($p < .05$) post-hoc pairwise comparisons, either successive iterations (collapsed across age) or between age groups (collapsed across iteration). (C) Integerness scores that measure the closeness of the responses to simple integers. The blue line corresponds to the baseline (shaded areas are one standard deviation obtained by bootstrapping 1,000 times) computed by comparing random locations to integer points (see Data Analysis and Statistics). Result shows responses are close to simple integer ratios well above chance.

To quantify the similarity of the three child groups to each other and to the adult group, we estimated Jenson-Shannon divergence (JSD) among groups and estimated the correlation between category weights among groups (see Table 2). The child group distributions did not differ significantly from one another ($p > .05$ in all cases) and all child groups' category weights were correlated with one another ($p < .001$ in all cases). All three child group distributions significantly differed from the Adult group distribution ($p < .05$ in all cases, uncorrected). When a Bonferoni correction was applied for multiple comparisons, only 6-7 year-old children significantly differed from Adults. All child group category weights were significantly correlated with the Adult group category weights ($p < .01$ in all cases). As expected, children displayed similar rhythm priors to adults from a comparable Western background. However, the differences among the probability distributions between the child

Groups	JSD p-value			Corr. of Weights: r (p-value)		
	1	2	3	1	2	3
			Age			
1: 6-7yrs						
2: 8-9yrs	0.065			.77 (<.001)		
3: 10-11yrs	0.25	0.29		.65 (.001)	.88 (<.001)	
4: Adults	<.001	0.014	0.03	0.56 (.006)	.78 (<.001)	0.76 (<.001)
			Music & Dance			
1: 0yrs						
2: 1-2yrs	0.16			0.89 (<.001)		
3: 3-8yrs	0.02	0.34		.76 (<.001)	0.87 (<.001)	
			Ethnicity & Region			
1: Hispanic (UNLV)						
2: Non-Hispanic (UNLV)	0.23			0.51 (0.016)		
3: Non-Hispanic (MAC)	0.029	0.32		.41 (.0594)	0.65 (0.001)	

Table 2. Comparisons among groups on Jenson-Shannon Divergence (JSD) and correlations among Category weights. Grouping variables: Age, Music & Dance Experience (max of either), and Ethnicity & Region (either Hispanic or Non-Hispanic, U: UNLV in Las Vegas, NV; M: McMaster University in London, ON). JSD p-value indicates the significance of each group comparison, where $p < .05$ indicates the distributions of the two groups significantly differ (uncorrected for multiple comparisons). The correlations among category weights indicates Pearson's r and the associated p-value, uncorrected.

groups and the adult group, as evidenced by JSD, suggests that the nature of these rhythm categories may be changing across childhood. Qualitatively, it is apparent that the distributions for the three child groups are less robust than in the Adult distribution for certain categories. Namely, while the modes at 1:1:2 and its rotations are quite prominent in all groups, the modes underlying other rhythm categories such as 3:3:2 appear much less robust for children than for adults. This suggests that perhaps the strength of these rhythm priors increases gradually throughout development. To get a more holistic understanding of group differences, we next examined category weights among the groups.

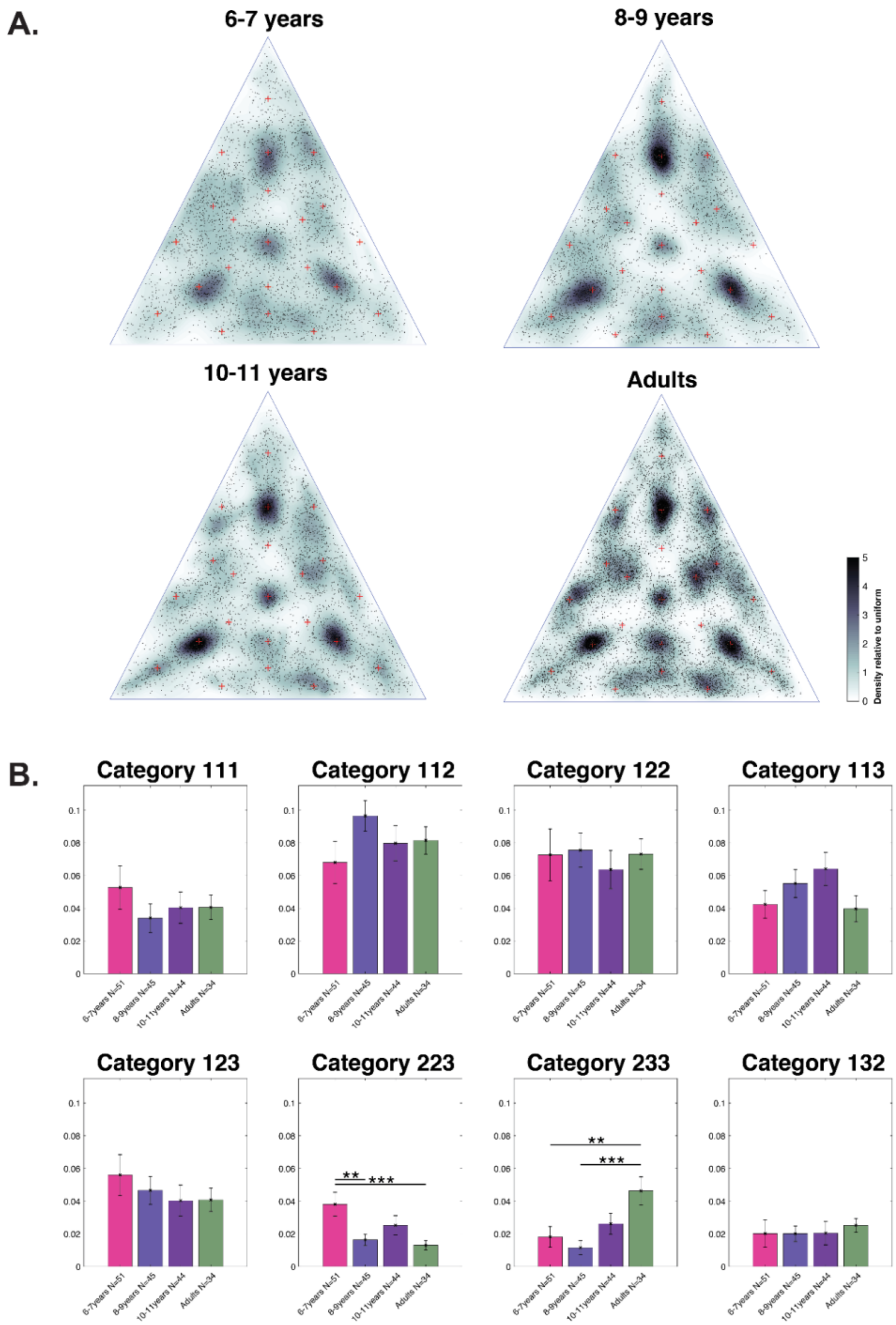

Figure 5. Results from the iterative tapping (IT) task. Note: The adult participants' data was originally published by Jacoby & McDermott (2017) in Experiment 1, used here with permission from the authors. (A) Kernel density estimation (KDE) of the continuous distribution underlying the data from iteration 5 by age group. Estimates are plotted on the rhythm simplex. Plus signs plot simple integer ratio rhythms. (B) Weights assigned to each rhythm category (collapsed across re-organizations of the same integer patterns, e.g. 1:1:2, 1:2:1, and 2:1:1 are collapsed into "Category 112" above). All statistical differences indicated are Bonferroni corrected for multiple comparisons: * $p < .05$, ** $p < .01$, *** $p < .001$.

3.1.3 Comparing Category Weights Across Age. The category weights for each of the eight unique rhythm categories (collapsing across rotations) are displayed in Figure 5B. Note that because each age group's category weights add up to 1, they must be interpreted collectively because they are not independent measurements. Two categories demonstrated significant group differences. Firstly, the category weights for 2:2:3 appears to decrease over age, with adults and 8- to 9-year-olds demonstrating a significantly lower category weight for 2:2:3 than 6- to 7-year-olds (Bonferroni-corrected). This rhythm category is not typical of Western music, and thus the underlying rhythm prior for North American listeners would be hypothesized to weaken over time due to lack of listening exposure to 2:2:3. Conversely, the category weight for 3:3:2 appears to increase with age, with Adults demonstrating a significantly higher category weight for 3:3:2 than 6-7 years or 8-9 years (Bonferroni-corrected). This rhythm category is relatively common in Western music (Biamonte, 2020; Toussaint, 2019), and thus the underlying rhythm prior for North American listeners would be hypothesized to increase over time. It is also worth noting that both rhythm categories are characteristic of more complex meters, in the sense that they do not have a common denominator. It is likely these category differences contribute to the significant JSD differences between the child groups and the Adult group, and it suggests enculturation may shape the strength of underlying rhythm priors well into the school age years, especially for the more complex integer ratios.

3.2 Rhythm Perception (RP) Task

Perception scores were submitted to a 2 (Rhythm Type: 4/4 vs. 7/8) x 3 (Age Group) mixed design ANOVA, which yielded a significant main effect of Rhythm Type ($F(1,91) = 7.874$, $p = .006$) and no effects of age group or interactions. As can be seen in Figure 6, perception scores were significantly higher for 4/4 meter than 7/8 meter ($t(92) = 3.03$, $p = .003$), suggesting participants at all ages were better at differentiating structure-disrupting from structure-preserving variations in the context of a familiar 4/4 meter with 2:1:1 rhythmic pattern than in

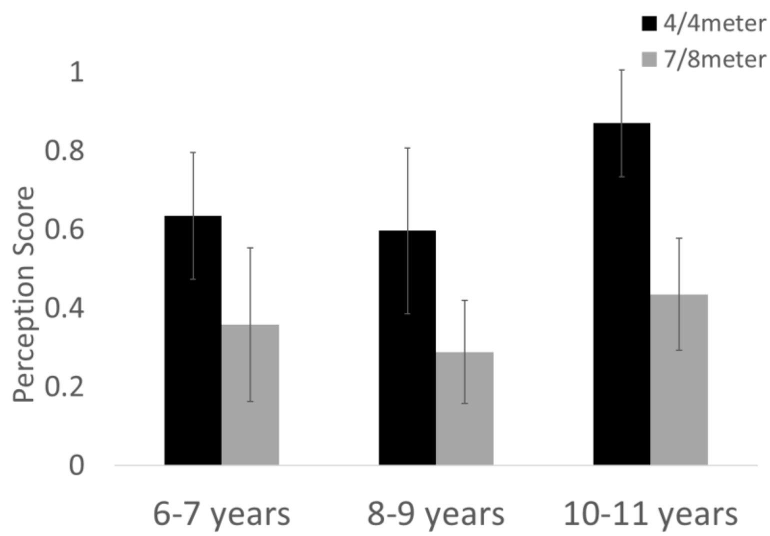

Figure 6. Results from the rhythm perception (RP) task, showing perception score (see methods) by age group. A positive perception score indicates greater similarity ratings were given to structure-preserving stimuli than to structure-disrupting stimuli. Black bars indicate 4/4 meter (culturally familiar) and gray bars indicate 7/8 meter (culturally-unfamiliar).

the context of an unfamiliar 7/8 meter with a 3:2:2 rhythmic pattern. Our results are comparable to the findings reported in Hannon et al. (2012) for similar age groups.

3.3 Tapping Baseline (TB) Task

Univariate ANOVAs on Spontaneous Motor Tempo, Tapping Variability, and Tapping Asynchrony in the Tapping Baseline (TB) task revealed no significant main effects of age group. Averages for the TB metrics are displayed in Table 3.

	Age 1	Age 2	Age 3
TB SMT	507.24 (183.80)	526.04 (197.63)	616.93 (278.89)
TB Asy	-18.57 (61.37)	-15.61 (58.99)	-35.07 (45.71)
TB CV	1.10 (0.44)	1.02 (0.45)	0.85 (0.42)

Table 3. Means and standard deviations by age group for the Tapping Baseline (TB) task. SMT = Spontaneous Motor Tempo. Asy= Asynchrony. CV = Coefficient of Variation.

3.3 Relation Between Rhythm Perception & Rhythm Production Tasks

Partial correlations (controlling for age) are displayed in Table 4. Results indicate a significant positive correlation between TB tapping coefficient of variation (CV) and overall IT tapping variability, suggesting baseline tapping to an isochronous metronome is relatively similar to the measure of tapping variability extracted from the iterative tapping task, where children tap to non-isochronous rhythms. IT 2:2:3 tapping variability is significantly correlated with both RP scores, producing a negative correlation with RP scores in 4/4 meter and a positive correlation with RP scores in 7/8 meter. Thus, greater accuracy perceptually discriminating songs with a 2:2:3 pattern in the RP task was associated with reduced variability when reproducing 2:2:3 rhythms in the IT task. Conversely, greater perceptually accuracy for the 2:1:1 pattern in the RP task was associated with increased variability for the 2:2:3 patterns in the IT task. Accuracy in the 4/4 RP condition was also negatively correlated with accuracy in the 7/8 RP condition. Together, these findings may suggest that with exposure to Western music, acquisition of a bias toward 2:1:1 rhythmic patterns is accompanied by a decrease in the bias towards less common 2:2:3 patterns.

3.4 Exploratory KDE Results Based on Musicianship and Regional Differences

To examine the potential effect of children's music or dance experience on rhythm priors, we examined the role of musicianship by creating 3 groups: 0 years, 1-2 years, and 3+ years of formal music or dance training (maximum of either). Note the three groups did not differ on the basis of age or SES ($p > .05$). We employed the same analyses used to compare age groups, but

	M	*SD*	1	2	3	4	5	6	7
1. BT tapping C.V.	1.01	0.44	-	.477***	.249*	.437*	.100	-.116	-.137
2. IT tap variability (all)	100.66	15.82	-	-	.654***	.756***	.319	.092	-.042
3. IT tap variability (1:1:2)	89.28	24.03	-	-	-	.042	.040	-.072	-.038
4. IT tap variability (2:3:3)	97.24	24.50	-	-	-	-	-.364	.123	.220
5. IT tap variability (2:2:3)	98.81	26.44	-	-	-	-	-	.588**	-.613**
6. RP Score (4-4 meter)	0.78	0.92	-	-	-	-	-	-	-.207*
7. RP Score (7-8 meter)	0.34	0.88	-	-	-	-	-	-	-

Table 4. Means, standard deviations, and partial correlations among variables for the Tapping Baseline (TB), Iterative Tapping (IT) and Rhythm Perception (RP) tasks. All partial correlations control for age. *: $p < .05$, **: $p < .01$, ***: $p < .001$.

this time with the musicianship groups, using JSD and category weights to estimate the similarity of the underlying KDE distributions and to correlate category weights among groups. While the 0 years and 1-2 years group distributions did not differ from one another ($p > .05$), the 0 years and 3-8 years group distributions did significantly differ from one another ($p < .05$). Category weights were strongly correlated among all three groups ($p < .001$ in all cases). Comparing groups on category weights revealed a significant difference for 1:2:3, such that 0 years showed a significantly higher category weight for 1:2:3 compared to the 3-8 years group ($p < .05$, Bonferroni corrected). All other category weight comparisons were not significant. Given the high degree of correlation among category weights, music and dance experience does not seem to have a large effect on *what* category weights are represented in the distribution, at least within a Western sample. However, differences in the overall distributions between non-musicians (0 years in both music and dance training) and musicians (3-8 years experience in either music or dance) suggest perhaps music experience changes the distribution of rhythm priors in other ways. Indeed, qualitative inspection of the distributions reveals a seemingly larger dispersal of distribution among categories for musicians (3-8 years) than both non-musicians (0 years) and novices (1-2 years) (see Supplemental Figure 2).

One future aim of this work is to examine cross-cultural differences in rhythm priors among children, especially in light of evidence for significant cross-cultural variation in rhythm priors related to region, ethnicity, and experience (university education, musicianship) recently reported by Jacoby and colleagues (Jacoby et al. 2021). At the same time, the current sample does offer the potential to explore this question informally within our sample due to the difference in population demographics among the testing locations. Specifically, we examined differences on the basis of two variables: Ethnicity and region. (UNLV – Las Vegas, NV, USA; McMaster – Hamilton, ON, Canada). Thus, we created three age-matched groups: Hispanic UNLV, Non-Hispanic UNLV, and Non-Hispanic McMaster. The three groups did not differ on the basis of age or SES ($p > .05$).

To quantify the similarity of the three groups to each other, we compared the groups' JSD measurements from each group's distribution to one another and the correlation of the integer ratio weights for each pair of groups. While the distributions did not differ between Hispanic (UNLV) children and Non-Hispanic (UNLV) children ($p > .05$), a significant difference in JSD was observed between Hispanic (UNLV) children and Non-Hispanic (McMaster) children ($p < .05$). Furthermore, while the category weights were significantly correlated between Hispanic

(UNLV) and Non-Hispanic (UNLV) and between Non-Hispanic (UNLV) and Non-Hispanic (McMaster) children, the correlation of category weights between Hispanic (UNLV) and Non-Hispanic (McMaster) children was not statistically significant (see Table 2). Category weights for 1:2:3 and 3:3:2 were larger in the Hispanic (UNLV) group than the Non-Hispanic (McMaster) group, and showed no difference compared to the Non-Hispanic (UNLV) group (see Supplemental Figure 3). This suggests an overall effect of region combined with ethnicity, such that the rhythm priors revealed by Hispanic children in Las Vegas are statistically different from the rhythm priors revealed by Non-Hispanic children from another region. These results suggest it is possible to reveal cultural differences in early childhood rhythm priors using this iterative paradigm. Moreover, because these differences relate to regional differences, this result corroborates prior work suggesting differences in rhythm priors are likely the result of differences in passive listening experience (Jacoby et al., 2021; Jacoby & McDermott, 2017). Of course, because this was determined from a sample of convenience, it is important for this effect to be replicated, with future work explicitly recruiting children from diverse regions and experiences.

4 General Discussion

This study provides the first measurement of rhythm priors in children over a broad rhythm space with no a priori hypotheses about the resulting distribution. All child age groups reached convergence by the final iteration, suggesting that, like adults, children's tapping behavior stabilizes on rhythm priors within five iterations, and supporting the idea that the paradigm can be used with younger samples. We found that in all children (6-11 years), the distribution of tapping responses suggested the presence of rhythm priors common in Western music and similar to those of adults with Western listening experience. Indeed, category weights were strongly correlated among all groups. However, at the same time, divergence in the distributions revealed significant differences among child and adult groups.

There are a few reasons why differences in the distributions between children and adults may have occurred. Firstly, children's tapping production skills are known to be less precise than those of adults, with children often tapping with much more variability and asynchrony than adults (McAuley et al., 2006; Monier & Droit-Volet, 2019). Our results also show significantly greater tapping variability in children compared to adults on the iterative task. Secondly, examining the Kernel Density Estimation (KDE) distributions from a qualitative lens reveals some apparent differences among the adult group and the child groups, with adults demonstrating apparently sharper peaks (visualized as warmer colors on the plots, see Figure 5) for some rhythm categories, such as 3:3:2, compared to children. It is possible this reflects less stability in children's rhythm priors compared to adults, although this needs to be tested empirically. This is not surprising given that many other temporal abilities continue to develop through childhood and adolescence, including musical beat and meter (Nave et al., in press; Nave-Blodgett et al., 2021b) and harmonic structure (Weiss et al., 2020). Taken together, our results suggest children's tapping behavior converges on culturally-familiar rhythm priors with simple integer ratios as early as 6 years of age, but the robustness of these priors may not reach adult-like levels until sometime after 10-11 years of age. Future work should test rhythm priors in other age ranges, including both younger children and adolescents, to obtain a full picture of the developmental trajectory of rhythm priors.

4.1 Parallels between Production and Perception

How does this unbound iterated paradigm compare to more common yet more limited measures of rhythm biases in children? To date, evidence for rhythm categories or priors in children has been obtained from relatively indirect perceptual tasks such as metronome matching or rhythm discrimination with a limited set of rhythm types. Data from such tasks have suggested both infants and children perform better on culturally-familiar, simple rhythms (e.g., 1:1:2 or 4/4 meter) compared to culturally-unfamiliar, complex rhythms (e.g., 2:2:3 or 7/8 meter) (Einarson & Trainor, 2016; Hannon & Trehub, 2005a, 2005b; Hannon et al., 2012). One aim of the present work was to validate the metrics obtained from the iterated paradigm against a perceptual paradigm previously used to estimate children's perceptual sensitivity to discrete rhythm categories. First, we found that tapping variability in the iterated task was highly correlated with variability in a more commonly used measure of children's production– synchronization with an isochronous metronome in the baseline tapping task. This suggests that the iterative tapping method, while optimized for group comparisons, can nevertheless be used to obtain a meaningful metric of individual production variability.

Second, we observed parallel patterns of performance across both perception and production tasks, providing converging evidence that children's rhythm categories already resemble adults' categories in important ways. Our perceptual task revealed a robust tendency at all ages for better performance discriminating rhythmic variation in songs with predominantly 2:1:1 and 1:1:2 ratios than a condition with songs having 2:2:3 and 3:2:2 ratios. This is consistent with prior work using this perceptual paradigm with both children (Hannon et al., 2012) and adults (Hannon & Trehub, 2005a; Hannon et al., 2011). Likewise, category weights were correlated among adults and all three child groups, and all groups showed higher weights for rhythm categories more common to Western music (like 2:1:1 and its rotations) than for categories less common (such as 2:2:3 and its rotations). Thus, both the perception and production tasks revealed a robust advantage for 2:1:1 rhythms over 2:2:3 rhythms across all age groups.

Our findings also provide evidence for developmental change. Even though children's category weights were correlated with adults, age-related changes were observed particularly for categories 2:2:3 and 3:3:2. The category weights for 3:3:2 and its rotations appeared to increase with age, while the weights for 2:2:3 and its rotations appeared to decrease with age. We also found that lower tapping variability on 2:2:3 rhythms was associated with a) higher rhythm perception scores in the 7/8 (2:2:3) condition and b) lower rhythm perception scores in the 4/4 (2:1:1) condition. These parallel findings between perception and production tasks support the interpretation that the age-related changes in category weights reflect real developmental changes. Individuals who are more variable (i.e. less stable) when reproducing the unfamiliar 2:2:3 rhythm are also worse at perceiving that rhythm, while they are better at perceiving the 2:1:1 rhythm. More broadly, this suggests that enculturation to familiar (Western) rhythms entails both stronger biases towards familiar categories and weaker biases toward unfamiliar categories.

It is worth noting, however, that this interpretation is limited because 1) the perceptual task provided no evidence that the bias towards the familiar 2:1:1 category increases with age, and 2) there were no correlations between perceptual performance and tapping variability for 2:1:1 or 3:3:2 categories. It is possible that because 1:1:2 is a universally robust category (see Jacoby et al. 2021), it emerges earlier in development than other categories. If this is the case, individual differences in 1:1:2 tapping variability in children might not provide meaningful variability that can be related to rhythm perception performance. As to why 3:3:2, a category that did show age-related change in the iterated task, did not predict perceptual performance might be related to the fact that the perceptual task did not directly measure discrimination of rhythms with the 3:3:2 structure. Future work should examine the perception of a wider range of discrete rhythm categories, including 3:3:2.

4.2 Effect of Experience on Children's Rhythm Priors

Prior work suggests that the iterative tapping paradigm is well-suited for revealing experience-dependent rhythm priors in adults. We explored the possibility that different listening experiences in our sample might result in divergent distributions. Our analysis based on musical experience revealed strongly correlated category weights among groups, although there was also significant divergence between distributions for non-musician (0 years music or dance) and musician (3-8 years music or dance) children. While this corroborates adult work demonstrating that musicians and non-musicians have similar rhythm priors with slightly different underlying distributions (Jacoby & McDermott, 2017), it is remarkable that a difference was observed among child musicians and non-musicians given the relatively low amount of formal music experience present in our sample. It should be noted, however, that deciding to train musically is correlated with other factors in the home environment such as parental encouragement (Dai & Shader, 2002), so it is possible that our results can be explained by something other than years of music training, and this should be further explored. Our analysis based on region and ethnicity revealed strongly correlated category weights among groups that shared a region (Hispanic-UNLV and Non-Hispanic-UNLV) and among groups from different regions but with similar ethnicity (Non-Hispanic-UNLV and Non-Hispanic-McMaster), but category weights were not correlated when groups differed on both region and ethnicity (Hispanic-UNLV and Non-Hispanic-McMaster). It is possible that this difference in rhythm prior distributions reflects differences in listening experience, presumably with children of self-reported Hispanic ethnicity having greater exposure to more Latin music traditions where rhythms prominently feature a 3:3:2 pattern (Toussaint, 2019). Of course, a more robust analysis of listening experience would utilize either self-reported type of music exposure as a grouping variable or more purposeful sampling procedures resulting in multiple diverse listening groups with diverse listening experiences. The present result provides a promising avenue for future work, aimed at understanding the developmental trajectory of differing listening experiences, incorporating multiple factors influencing experience such as music training, home musical environment and broader musical culture.

It is worth noting our sample was relatively high in socioeconomic status (SES), as estimated based on parent-reported highest level of education. Across groups, our participants' parents had an average of 4 years of college. Prior work suggests that college-educated listeners tend to under-represent cross-cultural diversity, with less variability among student groups compared to non-student groups, irrespective of cultural background (Jacoby et al., 2021). This suggests that in order to fully understand the scope of listener experience on rhythm priors, one must not only sample widely across different listening experiences on the basis of culture and ethnicity, but also on studentship and educational background. Because our sample is relatively highly educated, it is possible that the distributions underlying rhythm priors in children from less-educated backgrounds would differ to a greater extent.

4.4 Relation of Rhythm Categories to Musical Meter

Categorical rhythm priors can be assimilated to Western musical notation. However, whether categorical perception is sufficient for musical meter perception remains understudied.

Musical *meter* can be defined as multiple levels of co-occurring periodic structure (Hannon et al., 2018), which results in an alternating pattern of strong and weak beats (Lerdahl & Jackendoff, 1983). Some research has claimed that infants use meter to discriminate between rhythms (Hannon & Johnson, 2005; Phillips-Silver & Trainor, 2005; Winkler et al., 2009) while other research suggests that perception of musical meter is not fully developed until later in childhood and continues to develop throughout adolescence (Nave-Blodgett et al., 2021b). It is possible that categorical rhythm perception and meter perception, while related, rely on distinct processes with different developmental trajectories. Even though infants can discriminate rhythmic sequences that differ in musical meter, they may nevertheless rely more on rhythmic patterning or the ratios between intervals to succeed at the task. For example, a listener performing discrimination of rhythms with a 4/4 meter structure (1:1:2) can do so by paying attention to multiple levels of the beat (meter), or by continually comparing the stimulus to an internal prediction (based on a rhythm prior) that the underlying rhythm assimilates to a familiar rhythm category: 1:1:2. Importantly, this only requires the participant to pay attention to one level in the stimulus, that is the pattern of onsets, rather than two simultaneous levels as is implied by definitions of musical meter. It is possible that exposure to music gives rise to early emerging biases and preferences for culturally familiar rhythmic patterns (i.e., during infancy), but that young listeners cannot track multiple metrical levels of rhythmic structure until later in adolescence or adulthood (Ladinig et al., 2009; Nave-Blodgett et al., 2021b). Future work will need to disentangle categorical rhythm perception from meter perception by designing stimuli that can only be discriminated using either categorical mechanisms or hierarchical meter.

4.5 Conclusions

Many previous studies of the development of rhythm categories have relied on indirect measures with discrete sampling of a small subset of common rhythm structures. The present study serves as one of few direct measures of children's categorical rhythm perception and demonstrates the existence of rhythm priors in children that are consistent with their cultural listening experience. Our findings suggest that basic rhythm priors are likely acquired from passive experience prior to middle childhood, while more complex priors continue to be refined through childhood This study validates the use of the iterated tapping paradigm with younger populations and provides avenues for future research to further investigate rhythm priors during development, including the influence of type of experience (musical culture), nature of experience (informal or formal training), amount of experience, and maturation (age).

References

Anglada-Tort, M., Harrison, P., & Jacoby, N. (2022). REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. *Behavior research methods*, 1-15.

Biamonte, N. (2020). Rhythmic and Metric Theorisation in Rock Music. In A.F. Moore & P Carr (Eds.), *The Bloomsbury Handbook of Rock Music Research*. Bloomsbury.

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. https://doi.org/10.1177/0305735611425896

Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, *111*(40), 14559–14564. https://doi.org/10.1073/pnas.1406219111

Collier, G. L., & Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 602–627. https://doi.org/10.1037/0096-1523.21.3.602

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

Dai, D. Y., & Schader, R. M. (2002). Decisions regarding music training: Parental beliefs and values. *Gifted Child Quarterly*, *46*(2), 135-144.

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current Biology*, *31*(20), R1379–R1380. https://doi.org/10.1016/j.cub.2021.09.032

Dellatolas, G., Watier, L., Le Normand, M.-T., Lubart, T., & Chevrie-Muller, C. (2009). Rhythm Reproduction in Kindergarten, Reading Performance at Second Grade, and

Developmental Dyslexia Theories. *Archives of Clinical Neuropsychology*, *24*(6), 555–563. https://doi.org/10.1093/arclin/acp044

Desain, P., & Honing, H. (2003). The Formation of Rhythmic Categories and Metric Priming. *Perception*, *32*(3), 341–365. https://doi.org/10.1068/p3370

Einarson, K. M., & Trainor, L. J. (2016). Hearing the Beat: Young Children's Perceptual Sensitivity to Beat Alignment Varies According to Metric Structure. *Music Perception: An Interdisciplinary Journal*, *34*(1), 56–70. https://doi.org/10.1525/mp.2016.34.1.56

Essens, P. J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*(1), 1–7. https://doi.org/10.3758/BF03207132

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196

Fraisse, P. (1982). Rhythm and tempo. *The Psychology of Music*, *1*, 149–180.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & McAuley, J. D. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635–644. https://doi.org/10.1111/desc.12230

Hannon, E. E., Der Nederlanden, V. B., Christina, M., & Tichko, P. (2012). Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Annals of the New York Academy of Sciences*, *1252*(1), 92–99.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, *50*(4), 354–377. https://doi.org/10.1016/j.cogpsych.2004.09.003

Hannon, E. E., Nave-Blodgett, J. E., & Nave, K. M. (2018). The Developmental Origins of the Perception and Production of Musical Rhythm. *Child Development Perspectives*, *12*(3), 194–198. https://doi.org/10.1111/cdep.12285

Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science*, *14*(4), 865–872.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 543–548. https://doi.org/10.1037/a0027225

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.

Honing, H. (2013). 9—Structure and Interpretation of Rhythm in Music. In D. Deutsch (Ed.), *The Psychology of Music (Third Edition)* (pp. 369–404). Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00009-2

Jacoby, N., & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, *27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Jacoby, N., Polak, R., Grahn, J., Cameron, D., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., … McDermott, J. H. (2021). *Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors*. PsyArXiv. https://doi.org/10.31234/osf.io/b879v

Kalender, B., Trehub, S. E., & Schellenberg, E. G. (2013). Cross-cultural differences in meter perception. *Psychological Research*, *77*(2), 196–203. https://doi.org/10.1007/s00426-012-0427-y

Ladinig, O., Honing, H., Háden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception: An Interdisciplinary Journal*, *26*(4), 377–386.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception: An Interdisciplinary Journal*, *1*(2), 229–252.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter.* Oxford University Press.

Majtey, A. P., Lamberti, P. W., & Prato, D. P. (2005). Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. *Physical Review A*, *72*(5), 052310. https://doi.org/10.1103/PhysRevA.72.052310

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367. https://doi.org/10.1037/0096-3445.135.3.348

Monier, F., & Droit-Volet, S. (2019). Development of sensorimotor synchronization abilities: Motor and cognitive components. *Child Neuropsychology*, *25*(8), 1043–1062. https://doi.org/10.1080/09297049.2019.1569607

Murton, O., Zipse, L., Jacoby, N., & Shattuck-Hufnagel, S. (2017). Repetition and a Beat-Based Timing Framework: What Determines the Duration of Intervals Between Repetitions of a Tapping Pattern? *Timing & Time Perception*, *5*(3–4), 244–259. https://doi.org/10.1163/22134468-00002095

Nave, K. M., Snyder, J. S., & Hannon, E. E. (submitted). *Sustained musical beat perception develops into late childhood and predicts phonological abilities*.

Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021a). Auditory superiority for perceiving the beat level but not measure level in music. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(11), 1516–1542. https://doi.org/10.1037/xhp0000954

Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021b). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology: General*, *150*(2), 314–339. https://doi.org/10.1037/xge0000903

Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *Journal of Experimental Child Psychology*, *167*, 354–368. https://doi.org/10.1016/j.jecp.2017.11.009

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, *308*(5727), 1430–1430.

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic Prototypes Across CulturesA Comparative Study of Tapping Synchronization. *Music Perception*, *36*(1), 1–23. https://doi.org/10.1525/mp.2018.36.1.1

Povel, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 3.

Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Savini, T., Gamba, M. and Ravignani, A. (2023). Isochrony and rhythmic interaction in ape duetting. *Proceedings of the Royal Society B, 290*; 20222244.

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0007

Repp, B.H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969–992. https://doi.org/10.3758/BF03206433

Repp, B. H., London, J., & Keller, P. E. (2012a). Systematic distortions in musicians' reproduction of cyclic three-interval rhythms. *Music Perception: An Interdisciplinary Journal*, *30*(3), 291–305.

Repp, B. H., London, J., & Keller, P. E. (2012b). Distortions in Reproduction of Two-Interval Rhythms: When the "Attractor Ratio" Is Not Exactly 1:2. *Music Perception*, *30*(2), 205–223. https://doi.org/10.1525/mp.2012.30.2.205

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, *30*(18), 3544-3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: A cross-cultural comparison. *Developmental Psychology*, *46*(1), 286–292. https://doi.org/10.1037/a0017555

Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience. *PLOS ONE*, *10*(6), e0128839. https://doi.org/10.1371/journal.pone.0128839

Tierney, A., & Kraus, N. (2013). The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. *Journal of Neuroscience*, *33*(38), 14981–14988. https://doi.org/10.1523/JNEUROSCI.0612-13.2013

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231. https://doi.org/10.1016/j.bandl.2012.12.014

Toussaint, G.T. (2019). *The geometry of musical rhythm? What makes a "good" rhythm good?* CRC Press.

Ullal-Gupta, S., Hannon, E. E., & Snyder, J. S. (2014). Tapping to a Slow Tempo in the Presence of Simple and Complex Meters Reveals Experience-Specific Biases for Processing Music. *PLOS ONE*, *9*(7), e102962. https://doi.org/10.1371/journal.pone.0102962

Weiss, M. W., Cirelli, L. K., McDermott, J. H., & Trehub, S. E. (2020). Development of consonance preferences in Western listeners. *Journal of Experimental Psychology: General*, *149*(4), 634–649. https://doi.org/10.1037/xge0000680

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468–2471.

Wong, A. K. C., & You, M. (1985). Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-7*(5), 599–609. https://doi.org/10.1109/TPAMI.1985.4767707

Supplementary Tables and Figures.

	Age 1: 6-7 Years (N=56)		Age 2: 8-9 Years (N=50)		Age 3: 10-11 Years (N=52)	
	MAC[1]	UNLV	MAC	UNLV	MAC	UNLV
Sample Size[2]	n=26	n=30	n=20	n=30	n=5	n=47
TB Sample	n=14	n=29	n=20	n=27	n=5	n=42
RP Sample	n=0	n=28	n=0	n=25	n=0	n=42
WASI Sample	n=0	n=26	n=0	n=24	n=0	n=34
IT Sample	n=25	n=26	n=19	n=24	n=5	n=39
IT Num. Blocks (Excluded)	226 (24)	247 (13)	179 (11)	225 (15)	50 (0)	372 (18)
Mean Age	6.7±0.40	7.0±0.57	9.1±0.55	8.8±0.64	11.0±0.28	11.0±0.75
Females:Males	11:15	16:14	10:10	17:13	2:3	21:26
Music Experience[3]	1.00±1.55	1.07±1.31	2.70±2.62	0.73±1.11	2.60±2.30	1.47±2.03
SES[4]	4.77±0.99	4.96±1.45	5.10±1.07	4.38±1.63	5.4±0.89	4.93±1.15
WASI Composite Score	-	99.96±16.11	-	109.83±15.35	-	102.24±17.78

[1]Due to the COVID-19 pandemic, data collection was halted indefinitely for McMaster University. More data was collected at UNLV because local health regulations allowed them to open sooner.

[2]Music Experience scores were calculated as the maximum number of years of formal training in either music (instrument or voice) or dance.

[3]SES Scored on 6-point scaled based on highest level of parental education (1: No high school, 2: High school diploma, 3: Some college, 4: Technical School, 5: Four-year college/university degree, 6: Graduate degree).

Table S1. Descriptives for participants by age group, broken down by lab. Top: sample size broken down by age group, then further by testing location and task. TB= Tapping Baseline task. RP = Rhythm Perception task. WASI=Weschler's Abbreviate Scale of Intelligence. IT= Iterated Tapping task. IT num. blocks = Number of usable blocks (1 block = 1 seed rhythm with 5 tapped iterations), excluded blocks indicated in parentheses. Note that not all participants completed all tasks, as some attrition occurred before session 2 for some participants at UNLV. Bottom: descriptive data for each age group.

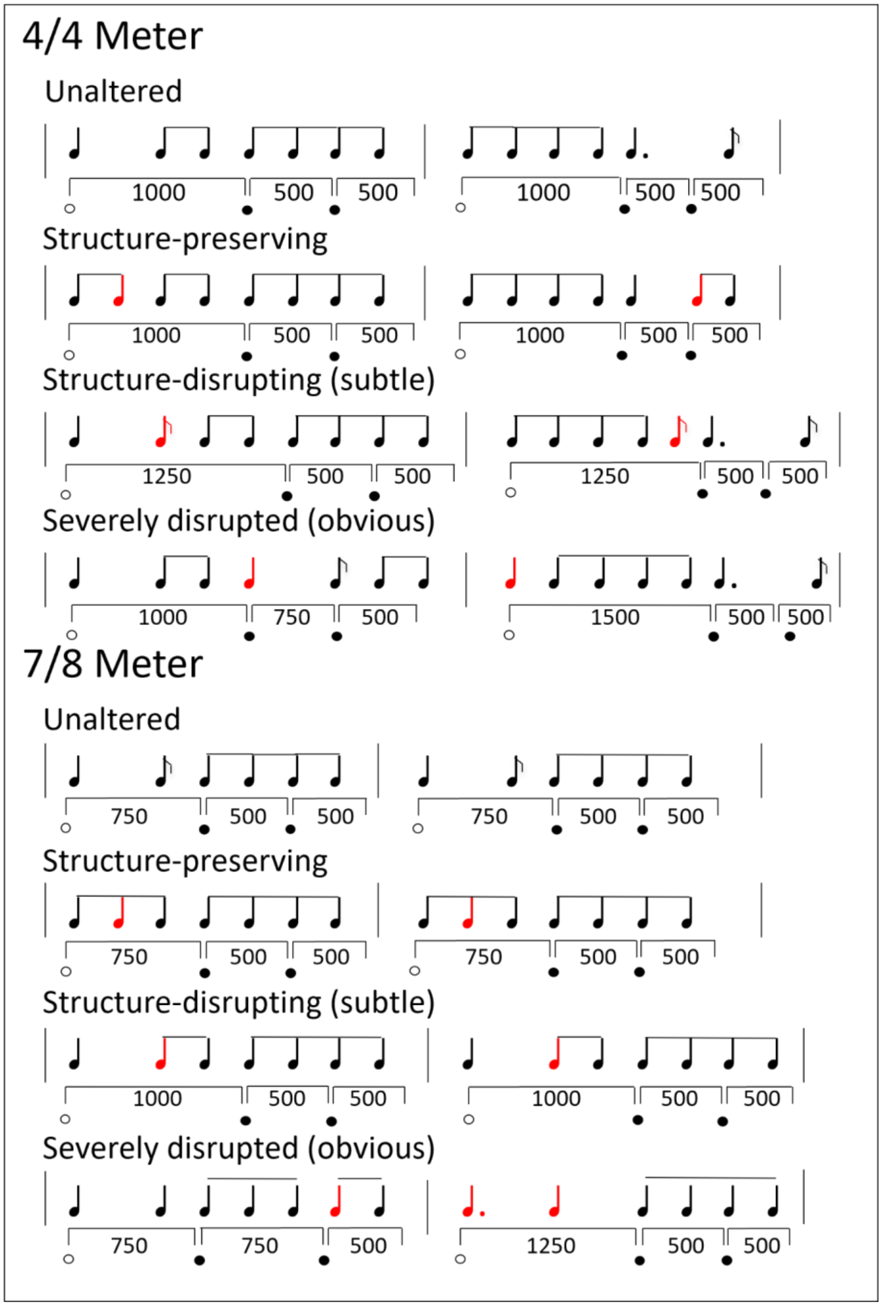

Figure S1. Examples of the stimuli used in the perception task (Animal Game). Stimuli included four types (unaltered, structure-preserving, structure-disrupting, and severely disrupted), which are depicted here for both the 4/4 meter (isochronous, culturally familiar) and 7/8 meter (non-isochronous, culturally unfamiliar). Added notes are depicted in red. Figure reproduced with permission from original authors, Hannon et al. (2012).

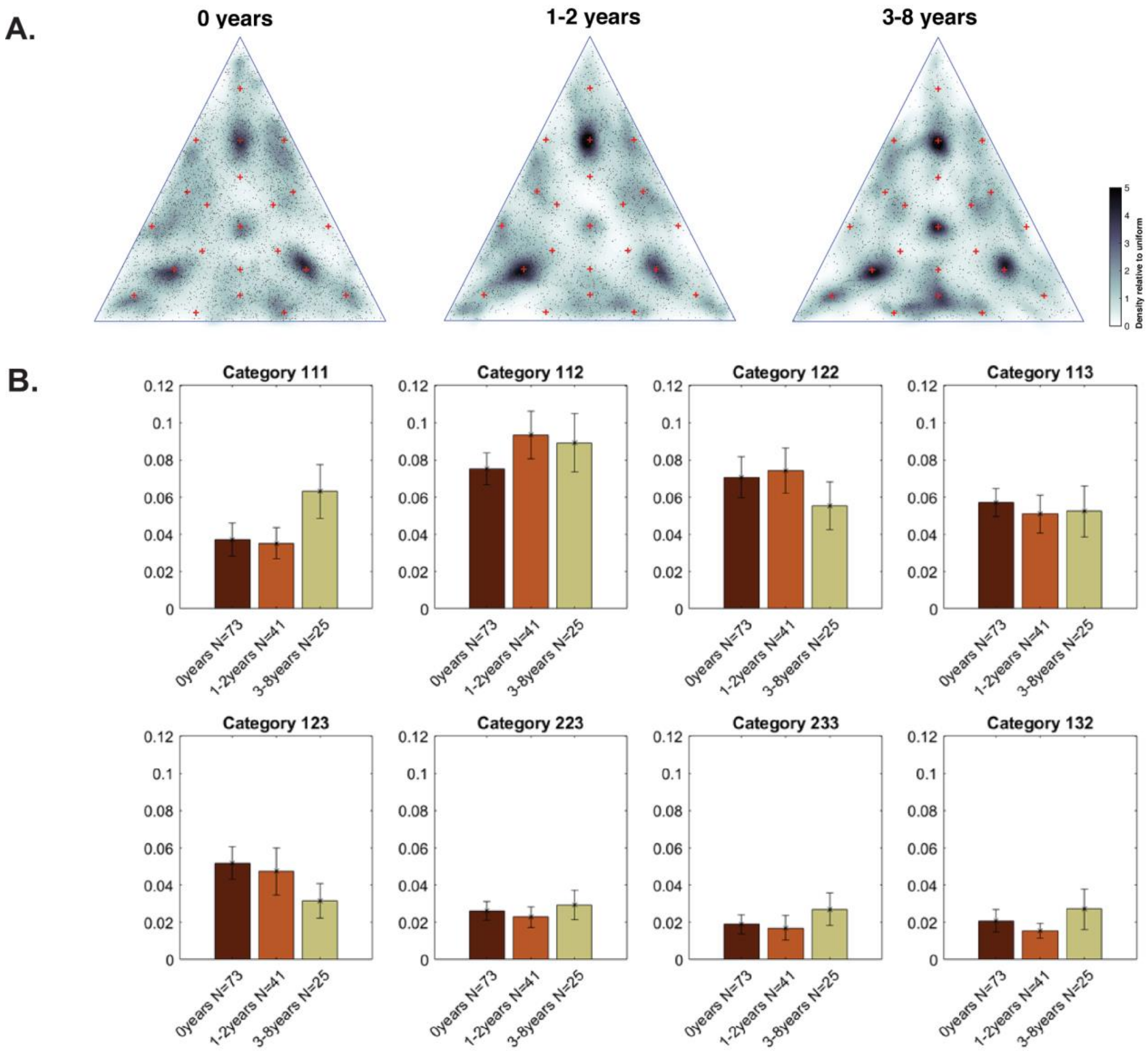

Figure S2. Results from IT task based on music and dance training. Top of figure shows Kernel density estimation (KDE) of the continuous distribution underlying the data from iteration 5 by music experience groups (0 years, 1-2 years, or 3-8 years). Estimates are plotted on the rhythm simplex. Crosses plot simple integer ratio rhythms. (B) weights assigned to each rhythm category (collapsed across re-organizations of the same integer patterns). All statistical differences indicated are Bonferroni corrected for multiple comparisons: * $p < .05$, ** $p < .01$, *** $p < .001$.

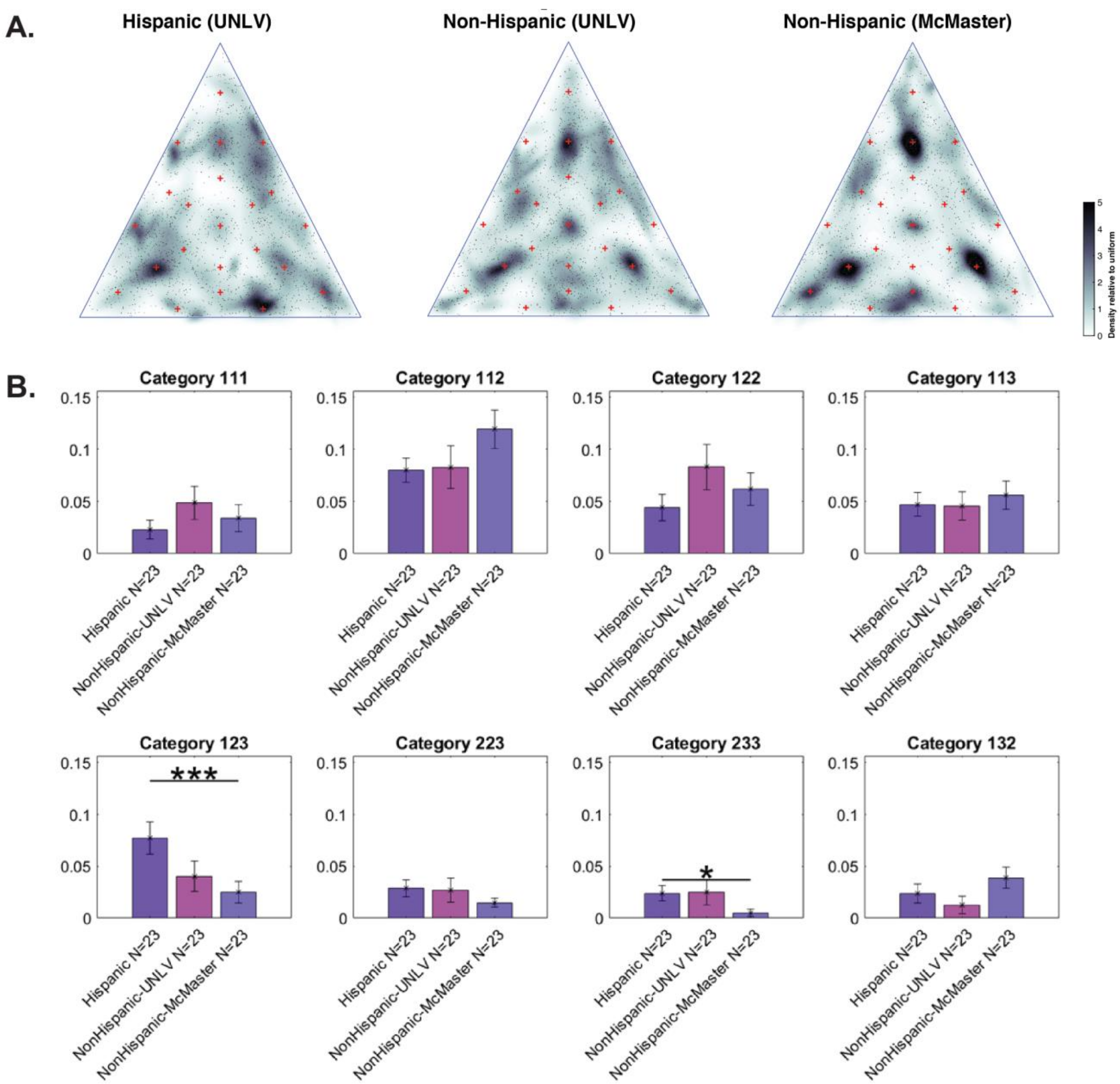

Figure S3. Results from IT task based on ethnicity and region. Top of figure shows Kernel density estimation (KDE) of the continuous distribution underlying the data from iteration 5 by ethnicity/region group (Hispanic (UNLV), Non-Hispanic (UNLV)< Non-Hispanic (McMaster). Note: UNLV is located in Las Vegas, Nevada, United States and McMaster is located in Hamilton, Ontario, Canada. Estimates are plotted on the rhythm simplex. Crosses plot simple integer ratio rhythms. (B) Weights assigned to each rhythm category (collapsed across re-organizations of the same integer patterns). All statistical differences indicated are Bonferroni corrected for multiple comparisons: * $p < .05$, ** $p < .01$, *** $p < .001$.