

Musical intervention enhances infants' neural processing of temporal structure in music and speech

T. Christina Zhao^{a,1} and Patricia K. Kuhl^{a,1}

^aInstitute for Learning & Brain Sciences, University of Washington, Seattle, WA 98195

Contributed by Patricia K. Kuhl, March 17, 2016 (sent for review November 8, 2015; reviewed by Andrea Chiba and Takako Fujioka)

Individuals with music training in early childhood show enhanced processing of musical sounds, an effect that generalizes to speech processing. However, the conclusions drawn from previous studies are limited due to the possible confounds of predisposition and other factors affecting musicians and nonmusicians. We used a randomized design to test the effects of a laboratory-controlled music intervention on young infants' neural processing of music and speech. Nine-month-old infants were randomly assigned to music (intervention) or play (control) activities for 12 sessions. The intervention targeted temporal structure learning using triple meter in music (e.g., waltz), which is difficult for infants, and it incorporated key characteristics of typical infant music classes to maximize learning (e.g., multimodal, social, and repetitive experiences). Controls had similar multimodal, social, repetitive play, but without music. Upon completion, infants' neural processing of temporal structure was tested in both music (tones in triple meter) and speech (foreign syllable structure). Infants' neural processing was quantified by the mismatch response (MMR) measured with a traditional oddball paradigm using magnetoencephalography (MEG). The intervention group exhibited significantly larger MMRs in response to music temporal structure violations in both auditory and prefrontal cortical regions. Identical results were obtained for temporal structure changes in speech. The intervention thus enhanced temporal structure processing not only in music, but also in speech, at 9 mo of age. We argue that the intervention enhanced infants' ability to extract temporal structure information and to predict future events in time, a skill affecting both music and speech processing.

infants | music | MEG | speech | early experience

Music training in early childhood has received increased attention as a model for the study of functional neural plasticity (1). Previous studies investigating musically trained adults and children have demonstrated their enhanced processing of musical pitch and meter in comparison with nontrained groups (2–6). Moreover, prior evidence also suggests generalization effects from early musical training to speech processing. For example, musically trained adults and children can better process pitch information in lexical tones and temporal information in syllable structure, compared with nonmusicians (7–10). These cross-domain effects from early music training to speech perception raise theoretically interesting and important questions about different levels of processing (e.g., lower level acoustic processing vs. higher level cognitive skills) affected by early experience (11).

However, there are several methodological issues preventing strong causal inferences about the effects of early music training in studies comparing musicians with nonmusicians. First, predispositions (e.g., higher auditory acuity) may lead individuals to self-select early music training, thus contributing to the observed differences between musicians and nonmusicians. Second, there exists great variability in the training received by musicians, including the nature, onset, and duration of musical training.

The current study combined three approaches to investigate the effects of early music experience: (i) We tested young infants using a randomized design, assigning them to either structured laboratory-controlled music intervention (“intervention”) or control activities (“control”). This approach allowed controlling

for effects related to predispositions (e.g., genetics) and prior music experience. (ii) We focused on temporal information processing such that the intervention targeted infants' learning of a specific meter (triple meter, e.g., the waltz) and tested the effects on both music (metrical structure) and speech (syllable structure). (iii) We used neural responses, measured by magnetoencephalography (MEG), as outcome measures to compare intervention and control infants in the spatial and temporal aspects of their cortical responses.

The primary goal of the current study was to investigate whether the intervention at 9 mo of age enhanced infants' neural processing of temporal structure in both music and speech. Our predictions followed the rationale that the intervention, targeting infants' learning of a specific meter, exerts influence at a higher level of processing. We argued that the intervention infants would become better at extracting the temporal pattern of complex sounds over time, leading to the ability to make more robust predictions of the timing of future stimuli based on the extracted temporal structure, an ability that would affect both music and speech processing. We predicted that, in the post-intervention/control MEG tests, the intervention group not only would process a learned temporal structure in music (i.e., triple meter) better than their control counterparts, but also would process a novel temporal structure in speech (i.e., a foreign syllable structure) better than controls.

We designed the current study (i.e., choice of age and number of intervention/control sessions), to parallel prior studies in this laboratory on infant speech learning at 9 mo of age (12, 13). This developmental stage constitutes a “sensitive period” for speech learning when infants' abilities to process speech can quickly change based on language experience (14, 15).

Significance

Musicians show enhanced musical pitch and meter processing, effects that generalize to speech. Yet potential differences between musicians and nonmusicians limit conclusions. We examined the effects of a randomized laboratory-controlled music intervention on music and speech processing in 9-mo-old infants. The Intervention exposed infants to music in triple meter (the waltz) in a social environment. Controls engaged in similar social play without music. After 12 sessions, infants' temporal information processing was assessed in music and speech using brain measures [magnetoencephalography (MEG)]. Compared with controls, intervention infants exhibited enhanced neural responses to temporal violations in both music and speech, in both auditory and prefrontal cortices. The intervention improves infants' detection and prediction of auditory patterns, skills important to music and speech.

Author contributions: T.C.Z. and P.K.K. designed research; T.C.Z. performed research; T.C.Z. analyzed data; and T.C.Z. and P.K.K. wrote the paper.

Reviewers: A.C., University of California, San Diego; and T.F., Stanford University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: zhaotc@uw.edu or pkkuhl@uw.edu.

Specifically, 47 9-mo-old infants raised in monolingual English-speaking environments with comparable prior and concurrent music listening experiences at home, whose parents were not musicians, were recruited (*Materials and Methods*). Infants were randomly assigned to the intervention or control group for 12 sessions (15 min each) of corresponding activity over a 4-wk period in the laboratory. The intervention sessions were designed to reflect naturalistic music training and to maximize infants' learning. The control sessions were designed to offer comparable visits to a laboratory, familiarity with the laboratory environment, levels of social interaction with other infants and caregivers, and levels of motor activity and engagement, but without music.

In the intervention sessions, infants experienced the triple meter (e.g., waltz) in various infant tunes and songs. Previous studies have demonstrated that infants at this age can rapidly learn temporal patterns in the music of their culture (16–18). We selected the triple meter (e.g., the waltz) because it has been demonstrated to be a more difficult temporal structure than duple meter (e.g., marching music) for infants at this age (19). We thus expected to see enhancement of triple meter processing due to intervention experience. Infants, with the aid of caregivers, tapped out the musical beats with maracas, or their feet, and were often bounced in synchronization to the musical beats, activities that are common in infant music classes (20). Control sessions had similar levels of social, physical activities. Infants, aided by their parents, played with toy cars, blocks, and other objects that required coordinated movements, such as moving and stacking, but without the musical component. In both the intervention and control sessions, infants were engaged in a social setting with one to two other infants and their caregivers, a setting demonstrated in previous work to be effective when infants are exposed to a foreign language (12). An experimenter facilitated each session by engaging the infants and their caregivers in the activities to a comparable degree.

To test whether the intervention enhanced infants' general ability to extract temporal structure and generate more robust predictions about future stimuli in complex auditory sounds, we examined their neural responses to temporal structure violations in both music and speech in temporal (auditory) as well as prefrontal cortical regions. The prefrontal region has been implicated in pattern processing and the predictive coding of auditory stimuli (21, 22). The mismatch response (MMR), measured with a traditional oddball paradigm within 2 wk of the last intervention/control session, was used to quantify neural processing. The magnitude of the MMR in the target cortical regions reflects neural sensitivity to the violation of temporal structure and thus the tracking and learning of that temporal structure (23). More specifically, in this paradigm, a standard stimulus is presented on ~85% of the trials to establish a temporal structure. A deviant stimulus violates this temporal structure and is randomly presented on the remaining 15% of the trials. Neural responses to all stimuli are recorded using magnetoencephalography (MEG), which measures the dynamic magnetic fields resulting from synchronized neural firing. The MMR is derived by first calculating a difference wave between neural responses to the standard stimuli and neural responses to the deviant stimuli; and it is generally characterized by a peak in amplitude in the difference wave between 150–250 ms after the onset of a change or violation in the auditory stimulus. The MMR is observed primarily in the temporal (auditory) regions of the cortex as well as the prefrontal regions, with a slightly delayed time course in the prefrontal cortex (24).

Traditionally, the MMR has been characterized using electroencephalography (EEG), which describes the response at the sensor level, in terms of its magnitude and polarity (i.e., negative vs. positive) referenced to a common sensor. Differences have been documented between infants and adults in the MMR with later peak latency, smaller magnitude, and a shift in polarity for infants from a positive to a negative MMR with age and

experience. The MMR has been considered fairly stable and readily observed across development (25, 26). MEG technology, with its excellent temporal resolution (millisecond) and good spatial resolution for measuring neural activities (27), allows examination of the MMR at the cortical level. Both the spatial and temporal patterns of brain activation, in both the prefrontal and temporal regions, can be examined. However, MEG uses different metrics to characterize the magnitude of neural response than EEG (*Materials and Methods, Source modeling*).

With MMR, we tested three specific hypotheses: (i) that the intervention group would exhibit a larger MMR response to violations in temporal structure for music compared with the control group, (ii) that the effects would be observed in both temporal (auditory) and prefrontal regions of the cortex, and (iii) that enhanced temporal structure processing, reflected by a larger MMR in temporal and prefrontal regions, would also be observed in response to speech syllable structure violation in the intervention group.

Results

To test the effects of the intervention on temporal structure processing in music (hypotheses *i* and *ii*), infants were presented with complex tones in triple meter structure in ~85% of the trials (group of three notes: strong–weak–weak). Occasionally (15% of the trials), the triple meter was violated through the removal of the last note in the group of three notes that constituted the triple meter (Fig. 1*A*) (details in *Results* and *Materials and Methods*). The strong notes immediately after the violations were deviants, and the strong notes before the violations were standards. Because the acoustic characteristics of standards and deviants are identical, any difference in infants' neural response therefore would reflect the detection only of temporal structure violation.

The neural responses to the standards and deviants were first preprocessed, averaged across trials, and projected from the MEG sensor space onto an infant cortical space using the

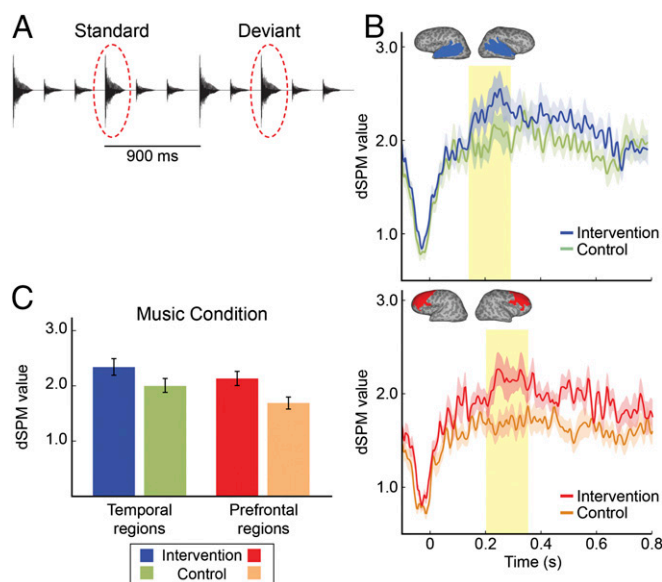


Fig. 1. Music condition (MEG). (*A*) Schematics of stimuli. Standard and deviant sounds are acoustically identical, and deviants violate the standard temporal structure. (*B, Top*) The group average of the difference waves for the temporal regions of the cortex for the intervention group and the control group. The shaded region indicates the selected time window for the MMR. Time 0 marks the onset of the strong beat. (*Bottom*) The group average of the difference waves for the prefrontal regions of the cortex for the intervention group and the control group. (*C*) Mean MMR values within the target time window by region (temporal region vs. prefrontal region) and group (intervention vs. control).

dynamic statistical parametric mapping (dSPM) method (28), resulting in statistically normalized values to characterize the neural activities (see *Materials and Methods*, *MEG individual analysis* for details). The difference waves were then calculated for each participant by subtracting neural responses to standards from deviants, and subsequently the magnitude of the differences was assessed, combining changes in both the strength and the direction of neural responses (*Materials and Methods*). The difference magnitudes in the temporal regions and prefrontal regions were further averaged for each participant. The target time window for the MMR in the temporal regions was selected as 150–300 ms postviolation and 200–350 ms postviolation for the prefrontal regions (Fig. 1*B*, shaded regions). These selections captured the peak of the response in the group average data and conformed to the classic time ranges for MMR documented in the infant literature (25, 29, 30). The MMRs in the target windows were then averaged for each participant.

The averaged values were submitted to a 2 (between group, intervention vs. control) \times 2 (within group, temporal regions vs. prefrontal regions) analysis-of-variance (ANOVA). The results revealed significant main effects for group [$F(1, 34) = 6.29, P = 0.017, \eta^2 = 0.16$] as well as for region [$F(1, 34) = 7.32, P = 0.011, \eta^2 = 0.18$] (Fig. 1*C*). No interaction between group and region was observed. These results support our first two hypotheses: The intervention group (mean = 2.23, SE = 0.11) exhibited larger MMR responses to temporal structure violations in the music condition compared with the control group (mean = 1.84, SE = 0.11), in both the auditory and prefrontal cortical regions.

Similarly, to test whether the intervention generalized to a new temporal structure in a new domain [speech (hypothesis *iii*)], the oddball paradigm was again used to measure infants' sensitivity to a violation in speech temporal structure (i.e., syllable structure). On 85% of the trials, infants were presented with a foreign syllable structure established using a disyllabic nonword with a long consonant between the vowels (i.e., /bibbi/); the syllable structure was violated by shortening the length of the middle consonant by 100 ms (i.e., /bibi/) (Fig. 2*A*, *Top*) (details in *Results* and *Materials and Methods*) in deviant trials occurring 15% of the time. This difference reflects an acoustic feature used in languages such as Japanese and Finnish, but not English (31). To achieve the identical statistical comparison for speech as in the music condition, wherein the responses to identical stimuli are compared while the stimuli occur in different contexts (e.g., as standard vs. as deviant), we adopted an established method (32) to record the neural response to /bibi/ when it was presented in a constant stream (as standard) in a separate short recording (Fig. 2*A*, *Bottom*). We subtracted neural responses to /bibi/ when it served as standard from neural responses to /bibi/ when it served as deviant in the context of the syllable /bibbi/. As in the case of music, the analysis window in both the temporal and the prefrontal regions was timed to the onset of the violation (onset of the second /bi/ syllable in /bibi/), which occurred 210 ms after the onset of the nonword (Fig. 2*B*, shaded region).

The same ANOVA model was used to address the hypothesis regarding the generalization of the effects to speech (Fig. 2*C*). A 2 (between group, intervention vs. control) \times 2 (within group, temporal regions vs. prefrontal regions) analysis was performed. As predicted, the results revealed a significant main effect of group [$F(1, 33) = 4.56, P = 0.039, \eta^2 = 0.12$] and of region [$F(1, 33) = 13.33, P = 0.001, \eta^2 = 0.29$]. No interaction between groups and regions was observed. Again, the intervention group (mean = 2.42, SE = 0.14) exhibited larger MMRs in response to temporal structure violations in speech compared with the control group (mean = 2.02, SE = 0.13). These effects occurred in both the auditory and prefrontal cortical regions, confirming our third hypothesis.

Discussion

The current study was designed to test three specific hypotheses: (i) that the 1-mo music intervention designed to help infants

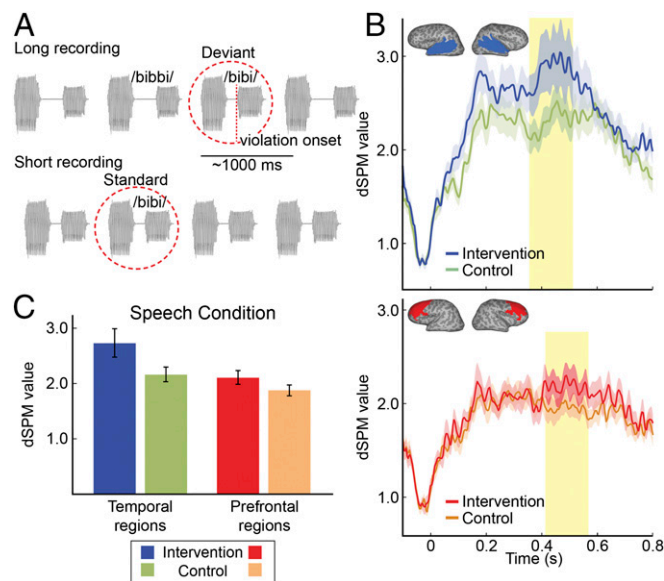


Fig. 2. Speech condition (MEG). (A) Schematics of stimuli. Deviants /bibi/ violate the syllable structure of /bibbi/. In a separate recording (*Bottom*), /bibi/ served as standards in a constant stream. (B, *Top*) The group average of the difference waves for the temporal regions of the cortex for the intervention group and the control group. The shaded region indicates the selected time window for the MMR, shifted accordingly with the onset of violation (210 ms after the onset of the nonword /bibi/, marked by time 0). (*Bottom*) The group average of the difference waves for the prefrontal regions of the cortex for the intervention group and the control group. (C) Mean MMR values within the target time window by region (temporal region vs. prefrontal region) and group (intervention vs. control).

learn a specific temporal structure in music (i.e., triple meter) would result in a larger neural response (MMR) in the intervention group to violations of temporal structure for music stimuli compared with the control group, (ii) that the effects would be observed in both temporal (auditory) and prefrontal regions of the infant cortex, and (iii) that enhanced temporal structure processing, reflected by a larger MMR in temporal and prefrontal regions, would also be observed in the intervention group when a completely new temporal structure was presented in the domain of speech. Our hypotheses were generated based on the rationale that the intervention group became better at extracting the temporal pattern of complex sounds and thus became more adept at predicting the timing of auditory stimuli based on the extracted temporal structure and that the ability of predictive coding is shared by both music and speech.

The results supported all three hypotheses. Our findings demonstrated that, as early as 9 mo of age, a randomized structured music intervention enhanced infants' neural processing of temporal structure in music, reflected by a significantly larger MMR in the intervention infants compared with the controls. As predicted, the effects were observed in both temporal and prefrontal cortical regions of the infant brain. Finally, the effects of the music intervention generalized to a new temporal structure change in a new domain, speech.

These results have implications for two long-standing issues in perception and suggest additional questions for future investigation: (i) the domain-specific vs. domain-general nature of music and speech processing, and (ii) infants' perception of patterns in complex sounds and the development of predictive coding.

The domain-specific vs. domain-general processing of complex sounds such as speech and music has been strongly debated (33, 34). Our current results provide data from the perspective that

across-domain generalization can occur as early as 9 mo of age from a music intervention to speech, during a period when infants are known to be undergoing an important transition in speech perception (14, 15). In the current study, we focused specifically on learning to extract higher level temporal information (i.e., temporal structure) from the intervention designed to simulate naturalistic music learning. Previous studies have suggested the significant role temporal information plays in speech perception and the impact of training using modified speech or nonspeech sounds to help infants and children prioritize specific temporal information, which may in turn enhance speech processing (35–38). However, the cross-domain generalization demonstrated here has not previously been tested or reported in young infants from music learning to speech processing.

Our results extend existing literature on within-domain effects from language experience to infants' speech processing during the sensitive period for speech learning. In previous studies, infants who experienced social foreign language intervention during this period learned to detect changes in foreign speech sounds better than controls who did not have such foreign language experience (12, 13). In the current study, we show that intervention in the music domain also affects foreign speech processing. In other words, our data suggest the possibility that the mechanisms supporting speech learning during this sensitive period are not exclusive to speech inputs; rather, a broader set of patterned auditory stimuli (e.g., music) can affect infants' speech processing. Future studies will be needed to replicate and extend this finding.

Secondly, our results have implications for the development of broader cognitive skills, such as the ability to detect patterns in sensory information. In our case, we examined the ability to extract temporal structure and to predict the timing of future stimuli. We predicted generalization effects from the intervention to speech based on the rationale that infants would learn to better attend to and extract auditory patterns in the temporal domain, allowing them to generate more robust predictions about the timing of future events based on learned patterns. Our results demonstrating enhanced foreign syllable structure processing in intervention infants strongly supports the idea that experience with music may enhance the development of a broader set of perceptual skills.

The ability to quickly extract patterns and predictively code future stimuli has been demonstrated in both adults and infants (21, 22, 39, 40), yet the potential that it may be enhanced through a music intervention in infancy is exciting. This idea corroborates recent evidence suggesting enhanced higher level cognitive abilities (e.g., working memory and executive functions) in musically trained adults and children (41–43). Future studies that specifically examine the relations between music learning in infancy and the development of cognitive skills (e.g., executive function) are warranted.

In addition, the current intervention generates many important questions for future research. We discuss one such question here concerning the involvement of other modalities (e.g., motor) in the development of auditory perception. Our intervention was designed to be maximally effective and to simulate important aspects of naturalistic music training for infants. We combined auditory experience with other modalities (e.g., motor) because it mirrors realistic infant music classes and supports the role of cross-modal coding that has been described as integral to music listening and learning (44–46). However, the exact contribution of the sensory–motor system in auditory learning was not targeted in the current study. Future studies are required to separate the effects of the perceptual and motor aspects of the intervention by developing additional control conditions that engage only the auditory system (e.g., passive listening intervention).

To summarize, the current study demonstrated that a music intervention designed for infants, incorporating key components of naturalistic early music training, enhanced infants' neural

processing of music temporal structure processing at 9 mo of age. Of equal importance, we observed robust generalization from the intervention to speech temporal structure processing. We interpret our results to suggest that the current 12-session music intervention at 9 mo of age may affect broad pattern extraction and predictive coding skills in young infants, skills shared by both music and speech processing. These results raise the possibility that enriched auditory environments, beyond enriched language experience, may be beneficial to infant learning.

Materials and Methods

Participants. Forty-seven infants born and raised in monolingual English-speaking families were recruited at 40 wk of age. The inclusion criteria included the following: (i) full term and born within 14 d of due date, (ii) no known health problems and no more than three ear infections, (iii) birth weight ranging from 6 lb to 10 lb, and (iv) no previous or concurrent enrollment in infant music classes. Experimental procedures were approved by the Institute Review Board of the University of Washington, and all informed consents were obtained from the parents of the infants.

Infants were randomly assigned to either the intervention group or the control group. Questionnaires filled out by the parents ensured that the two groups experienced comparable music listening in their home environments (intervention, 9.93 ± 6.83 h/wk; control, 12.89 ± 9.47 h/wk, $t(36) = -1.1$, $P = 0.28$). Participation required completion of 12 intervention or control sessions over a 4-wk period, and up to three MEG recordings to ensure completion of tests on both music and speech conditions within 2 wk of the last intervention or control session. Overall, one infant failed to complete all intervention sessions and seven failed to complete MEG recordings due to fussiness. The final sample of infants who completed all 12 intervention/control sessions, as well as the MEG test sessions was as follows: intervention group ($n = 20$) and control group ($n = 19$). In addition, three MEG recordings from the music condition and eight from the speech condition failed to produce usable data due to the following: excessive movement (*MEG preprocessing*) (two recordings), too few usable trials (two recordings), and technical failure (seven recordings). For the music condition, MEG recordings from 36 participants were included in analysis (18 from intervention, 12 male; 18 from control, 9 male). For the speech condition, MEG recordings from 35 participants were included in analysis of the speech condition (16 from intervention, 12 male; 19 from control, 9 male). Infants with successful MEG recordings were further recruited to complete a structural MRI scan within 2 wk of the last MEG recording. An MRI scan from one subject was obtained successfully and was used to construct the head model.

Stimuli.

Intervention/control phase. For the intervention group, recordings of children's music in triple meter were selected from various commercially published music CDs for infants and toddlers. They were selected to vary in tempo (slow to fast; range, 115–180 beats per minute) and voices (for songs) to facilitate the learning and extraction of the abstract temporal structure. All music was recorded on six CDs of about 15 min duration.

MEG testing phase.

Music condition. The triple meter structure was created by combining a strong complex tone with two weak complex tones with sound-onset-asynchrony (SOA) of 300 ms. The strong tone was created by amplifying the weak tone by 10 dB in Audacity software (version 2.0; Sound Forge). The complex tone (duration, 200 ms; sampling frequency, 44.1 kHz) had a fundamental frequency of 220 Hz (A3) and was synthesized by combining a tone with "grand piano" timbre with a woodblock sound in Overture software (version 4; Sonic Scores). In total, there were 1,250 trials, with 200 deviant trials.

Speech condition. The disyllabic nonword stimuli were created in Praat software by combining a synthesized syllable /bi/ with silent gaps in between (47). The syllable /bi/ was synthesized (duration, 160 ms; sampling frequency, 44.1 kHz; fundamental frequency, 220 Hz) to have 30 ms of formant transition at the beginning and at the end, as well as 100 ms of steady-state vowel. The disyllabic nonword /bibbi/ was created by combining two syllables with 150 ms of silence in between, and /bibi/ was created by reducing the duration of the silence to 50 ms. For both stimuli, the first syllable was amplified by 5 dB to create a strong–weak stress pattern.

Separate stimulus sequences were created for the two recordings. In a long recording, 1,250 trials were played of which 200 were deviants (/bibbi/). In a short recording, 200 trials of stimulus /bibi/ were played (Fig. 2A, *Bottom*). The SOAs were jittered between 900 ms and 1,100 ms to minimize effects associated with predictability of the onset of the first syllable (Fig. 2A, *Top*). This procedure ensured that infants extracted the temporal structure of the

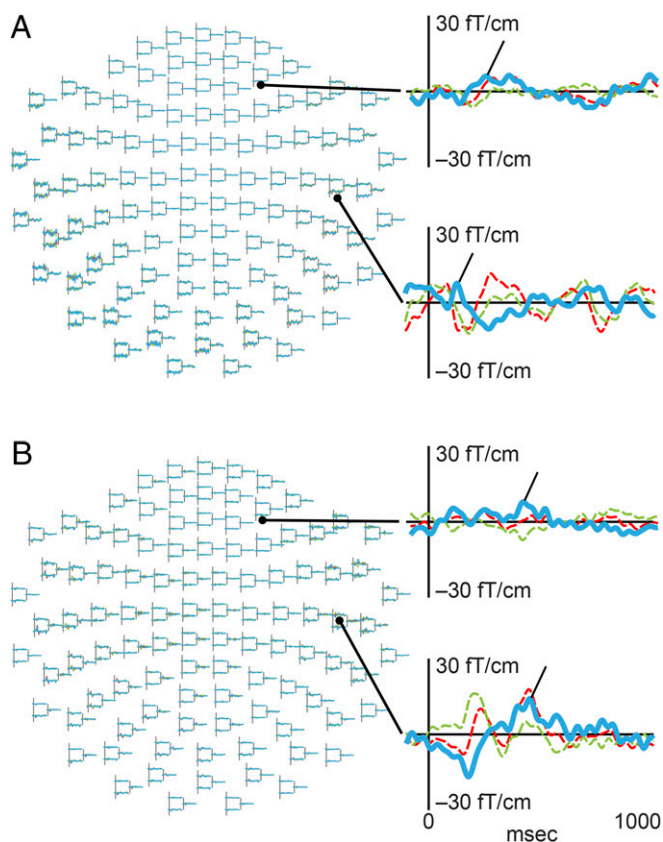


Fig. 3. (A) Music condition (sensor data from one participant). Red line, averaged epochs for standards; green line, average epochs for deviants; blue line, difference between standards and deviants. Two channels were selected to illustrate responses to the standards and deviants as well as the difference waves in the temporal and frontal areas at the sensor level. (B) Speech condition (sensor data from one participant). Red line, averaged epochs for /bibi/, serving as standards; green line, average epochs for /bibi/ deviants; blue line, difference between /bibi/ serving as standards and deviants. Two channels were selected to illustrate responses to the standards and deviants as well as the difference waves in the temporal and frontal areas at the sensor level.

standard stimulus intersyllabically, not by merely tracking the stimulus onset at a set interval.

Equipment and Procedure.

Intervention phase.

Intervention group. Infants assigned to the intervention group completed 12 sessions (15 min per session) of structured music intervention over a 4-wk period. This protocol design was in line with previous studies examining foreign language intervention in this age range, with consideration of practicalities such as caregivers' availability and the duration of time infants can stay attentive without being fussy. The sessions took place in a sound-attenuating booth decorated to be infant friendly. In each session, one of the six CDs was played through two speakers at a comfortable listening level of 65 decibels (A-weighted sound levels) (dBA), measured at the center of the room. Four video cameras were placed at different locations in the room to capture the behaviors of the infants during all sessions. Up to three infants and their primary caregivers were in the room, along with an experimenter who facilitated the session. The caregivers were instructed to interact with the infant throughout the sessions, with the aim of synchronizing the infants' movements to the musical beats. A variety of infant-safe simple percussive musical toys were introduced to infants to facilitate infants' movements, such as shaking maracas, and foot tapping and bouncing were also used.

Control group. Infants assigned to the control group completed 12 sessions of social free play with nonmusical toys appropriate to the infants' age. The sessions took place in the same sound-attenuating booth, decorated to be

infant friendly, used for the intervention group of infants. In each session, up to three infants and their primary caregivers were in the room, along with an experimenter. The infants were engaged in activities with the caregivers, other infants, and the experimenter to a degree comparable with the intervention group through the introduction of various nonmusical toys. **MEG testing phase.** Infants completed their MEG recordings within 2 wk of the last intervention/control session. The order of testing for speech and music was counterbalanced across infants.

Stimulus presentation. Auditory stimuli used in the tests were delivered using the Psychophysics Toolbox in MATLAB (48) on an HP workstation connected to TDT RP 2.7 hardware (Tucker-Davis Technologies hardware). All stimuli were processed such that their rms values were referenced to 0.01, and they were further resampled to 24,414 Hz for the TDT. Subsequently, the sounds were played through a speaker with a flat frequency response at a comfortable listening level of 65 dBA, measured under the MEG dewar.

MEG measurement. All MEG data were acquired inside a magnetically shielded room (MSR) (IMEDCO) using a MEG (306-channel Elekta Neromag) system with 204 planar gradiometers and 102 magnetometers. All data were acquired at a 1-kHz sampling frequency.

In a typical MEG session, the infant was first seated in a customized high chair outside of the MSR. A research assistant distracted the infants while the technician fit a stretch cap on infants' heads. One pair of electro-oculogram (EOG) electrodes was attached to the lower corner of the left eye and upper corner of the right eye to measure eye blinks. Five head position indicator (HPI) coils were attached to the cap to measure head position continuously under the MEG dewar. Three landmarks (left preauricular point, right preauricular point, and nasion) and the five HPI coils were digitized along with 100 additional points along the head surface with an electromagnetic 3D digitizer (Fastrak; Polhemus). Then the infant was placed under the MEG dewar in a customized chair. A research assistant continued to distract the infant with toys, and the primary caregiver was seated next to the MEG machine. Once the infant seemed to be calm and alert, the MEG recording started and the stimulus presentation began.

In addition, at the end of each MEG session, a 5-min empty-room recording was made with the same stimuli playing.

MRI structural scan. The MRI structural scans were completed within 2 wk after the last MEG session using a 3.0T system with an eight-channel head coil (Achieva; Phillips). A multiecho T1 pulse sequence (3D water excited/Turbo field echo) was used with the following parameters: repetition time (TR), 24 ms; inversion time (TI), 1,450 ms; and echo times (TEs), 6.5 ms, 12.2 ms, and 18 ms; acquisition voxel size, 0.37 mm³; sensitivity encoding (SENSE) factor, 2.5 in the anterior–posterior direction.

Data Analysis.

Head model template creation. An MRI scan obtained from one participant was used to create the template head model. The images were first processed by calculating the root-mean-square (rms) of the values obtained from the three echoes for each voxel. The resulting images were segmented in FMRIB Software Library-FMRIB's Automated Segmentation Tool (FSL-FAST) (49). The white matter component resulting from the segmentation was then used to process the images again to enhance the signal for the white matter. Cortical reconstruction and volumetric segmentation were performed using the FreeSurfer image analysis suite (surfer.nmr.mgh.harvard.edu). A surface-based cortical source space was created using the topology of a recursively subdivided icosahedron 5, resulting in ~20,484 source points distributed throughout cortical surfaces. In addition, a subcortical volumetric source space with grid spacing of 5 mm was constructed, including ~4,425 source points distributed throughout subcortical structures and the cerebellum.

MEG preprocessing. The raw MEG recordings underwent a series of standardized preprocessing steps for noise suppression. The temporal signal space separation (tSSS) and head movement compensation aligning the data to the mean head position were used first (Elekta MaxFilter 2.2) to suppress noise from outside of the MEG dewar and to compensate for effects related to infants' head movement during the recording. This procedure was designed to improve the signal-to-noise ratio of the data by suppressing external interference (i.e., noise from outside of the helmet) without introducing excessive reconstruction noise (50, 51). The infant head movement was evaluated by assessing the maximum SD of the center head position across all time points. Then, the signal-space projection (SSP) method was adopted to isolate components of physiological artifacts (i.e., heartbeats and eye blinks), using in-house MATLAB scripts (52). Lastly, the signal was band-pass filtered from 1 to 40 Hz, and noisy and dead channels were rejected based on the overall power calculated of each channel.

MEG individual analysis.

Epoch average. Epochs were rejected when the peak-to-peak amplitude was over 1.5 pT/cm for gradiometers or 2.0 pT/cm for magnetometers. Epochs

(–50 to 900 ms) in response to standards and deviants were then averaged separately for each subject after baseline correction. Baseline correction was accomplished by subtracting the mean value of the time period before trial onset (–50 to 0 ms) from the epoch. Data from one exemplar subject at the sensor level is demonstrated from the music condition (Fig. 3A) and from the speech condition (Fig. 3B).

Source modeling. Forward modeling used the boundary element method (BEM) isolated-skull approach with inner skull surface extracted from the MRI of the template. Both the source space and the BEM surface were then aligned and scaled to optimally fit each subject's head shape revealed by head digitization points. All modeling was done with in-house MATLAB scripts in combination with the MNE software suite (53).

Inverse source modeling was performed using the dynamic statistical parametric mapping (dSPM) method without dipole orientation constraints and with data from both gradiometers and magnetometers (28). The source activities were normalized to the noise covariance computed from the corresponding empty-room recording, which underwent the same preprocessing steps except for the movement compensation. This procedure resulted in statistically normalized scores for three dipole components at each source location for each time point (i.e., dipole strengths in three orthogonal directions). The difference between standards and deviants was then computed

for each source location at each time point through the following: (i) subtraction in each of the dipole components and (ii) calculating the magnitude of the difference wave (hereafter, difference magnitude). Computation of the difference between standards and deviants takes into consideration both dipole strength and direction at each source location such that the magnitude value combines changes in both dimensions.

Group comparison. The difference magnitudes of each subject were interpolated onto a spherical atlas for group level inferences. The FreeSurfer Destrieux atlas was also projected onto this spherical atlas for labeling each source point. Based on the FreeSurfer labeling, difference magnitudes in the temporal regions and prefrontal regions were then averaged separately for each subject. The prefrontal regions included superior, middle, and inferior gyri and sulci of the frontal lobe; the temporal regions included the superior and middle gyri and sulci of the temporal lobes. The brain region selected for prefrontal analysis was broad given the use of one infant head template instead of individual MRIs for all infants.

ACKNOWLEDGMENTS. The research described here was supported by National Science Foundation Science of Learning Center Program Grant SMA-0835854 to the University of Washington (UW) LIFE Center (P.K.K., principal investigator), the Ready Mind Project at the UW Institute for Learning & Brain Sciences, and a grant from the Washington State Life Sciences Discovery Fund (LSDF).

- Zatorre RJ (2013) Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science* 342(6158):585–589.
- Koelsch S, Schröger E, Tervaniemi M (1999) Superior pre-attentive auditory processing in musicians. *Neuroreport* 10(6):1309–1313.
- Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain* 129(Pt 10):2593–2608.
- Pantev C, et al. (1998) Increased auditory cortical representation in musicians. *Nature* 392(6678):811–814.
- Vuust P, et al. (2005) To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24(2):560–564.
- Geiser E, Sandmann P, Jäncke L, Meyer M (2010) Refinement of metre perception: Training increases hierarchical metre processing. *Eur J Neurosci* 32(11):1979–1985.
- Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10(4):420–422.
- Marques C, Moreno S, Castro SL, Besson M (2007) Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *J Cogn Neurosci* 19(9):1453–1463.
- Magne C, Schön D, Besson M (2006) Musicians children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *J Cogn Neurosci* 18(2):199–211.
- Marie C, Magne C, Besson M (2011) Musicians and the metric structure of words. *J Cogn Neurosci* 23(2):294–305.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11(8):599–605.
- Kuhl PK, Tsao FM, Liu HM (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 100(15):9096–9101.
- Conboy BT, Kuhl PK (2011) Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Dev Sci* 14(2):242–248.
- Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7(1):49–63.
- Kuhl PK, et al. (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci* 9(2):F13–F21.
- Hannon EE, Trehub SE (2005) Metrical categories in infancy and adulthood. *Psychol Sci* 16(1):48–55.
- Hannon EE, Trehub SE (2005) Tuning in to musical rhythms: Infants learn more readily than adults. *Proc Natl Acad Sci USA* 102(35):12639–12643.
- Gerry DW, Faux AL, Trainor LJ (2010) Effects of Kindermusik training on infants' rhythmic enculturation. *Dev Sci* 13(3):545–551.
- Bergeson TR, Trehub SE (2006) Infants' perception of rhythmic patterns. *Music Percept* 23(4):345–360.
- Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308(5727):1430.
- Schwartz M, Kotz SA (2013) A dual-pathway neural architecture for specific temporal prediction. *Neurosci Biobehav Rev* 37(10 Pt 2):2587–2596.
- Bekinschtein TA, et al. (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci USA* 106(5):1672–1677.
- Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends Cogn Sci* 13(12):532–540.
- Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Näätänen R (2000) Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage* 12(1):14–19.
- Cheour M, Leppänen PHT, Kraus N (2000) Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clin Neurophysiol* 111(1):4–16.
- Morr ML, Shafer VL, Kreuzer JA, Kurtzberg D (2002) Maturation of mismatch negativity in typically developing infants and preschool children. *Ear Hear* 23(2):118–136.
- Hamalainen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65(2):413–497.
- Dale AM, et al. (2000) Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26(1):55–67.
- Cheour M, et al. (1998) Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1(5):351–353.
- Kuhl PK, Ramirez RR, Bosseler A, Lin J-FL, Imada T (2014) Infants' brain responses to speech suggest analysis by synthesis. *Proc Natl Acad Sci USA* 111(31):11238–11245.
- Aoyama K (2001) A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions (Kluwer Academic, Boston).
- Kujala T, Kallio J, Tervaniemi M, Näätänen R (2001) The mismatch negativity as an index of temporal processing in audition. *Clin Neurophysiol* 112(9):1712–1719.
- Patel AD (2014) Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear Res* 308:98–108.
- Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front Psychol* 2:142.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270(5234):303–304.
- Tallal P, et al. (1996) Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* 271(5245):81–84.
- Merzenich MM, et al. (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 271(5245):77–81.
- Benasich AA, Choudhury NA, Realpe-Bonilla T, Roesler CP (2014) Plasticity in developing brain: Active auditory exposure impacts prelinguistic acoustic mapping. *J Neurosci* 34(40):13349–13363.
- Basirat A, Dehaene S, Dehaene-Lambertz G (2014) A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition* 132(2):137–150.
- Emberson LL, Richards JE, Aslin RN (2015) Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proc Natl Acad Sci USA* 112(31):9585–9590.
- Moreno S, et al. (2011) Short-term music training enhances verbal intelligence and executive function. *Psychol Sci* 22(11):1425–1433.
- Zuk J, Benjamin C, Kenyon A, Gaab N (2014) Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One* 9(6):e99868.
- Kraus N, Strait DL, Parbery-Clark A (2012) Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Ann N Y Acad Sci* 1252:100–107.
- Khalil AK, Minces V, McLoughlin G, Chiba A (2013) Group rhythmic synchrony and attention in children. *Front Psychol* 4:564.
- Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front Syst Neurosci* 8:57.
- Fujioka T, Fidalci BC, Ross B (2014) Neural correlates of intentional switching from ternary to binary meter in a musical hemiola pattern. *Front Psychol* 5:1257.
- Boersma P, Weenink D (2001) Praat: A system for doing phonetics by computer. *Glott Int* 5(9/10):341–345.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20(1):45–57.
- Taulu S, Hari R (2009) Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp* 30(5):1524–1534.
- Taulu S, Kajola M (2005) Presentation of electromagnetic multichannel data: The signal space separation method. *J Appl Phys* 97(12):124905.
- Uusitalo MA, Ilmoniemi RJ (1997) Signal-space projection method for separating MEG or EEG into components. *Med Biol Eng Comput* 35(2):135–140.
- Gramfort A, et al. (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86:446–460.