

Efficacy of Auditory versus Motor Learning for Skilled and Novice Performers

Rachel M. Brown* and Virginia B. Penhune

Abstract

■ Humans must learn a variety of sensorimotor skills, yet the relative contributions of sensory and motor information to skill acquisition remain unclear. Here we compare the behavioral and neural contributions of perceptual learning to that of motor learning, and we test whether these contributions depend on the expertise of the learner. Pianists and nonmusicians learned to perform novel melodies on a piano during fMRI scanning in four learning conditions: listening (auditory learning), performing without auditory feedback (motor learning), performing with auditory feedback (auditory–motor learning), or observing visual cues without performing or listening (cue-only learning). Visual cues were present in every learning condition and consisted of musical notation for pianists and spatial cues for nonmusicians. Melodies were performed from memory with no visual cues and with auditory feedback (recall) five times during learning. Pianists showed greater improvements in pitch and

rhythm accuracy at recall during auditory learning compared with motor learning. Nonmusicians demonstrated greater rhythm improvements at recall during auditory learning compared with all other learning conditions. Pianists showed greater primary motor response at recall during auditory learning compared with motor learning, and response in this region during auditory learning correlated with pitch accuracy at recall and with auditory–premotor network response during auditory learning. Nonmusicians showed greater inferior parietal response during auditory compared with auditory–motor learning, and response in this region correlated with pitch accuracy at recall. Results suggest an advantage for perceptual learning compared with motor learning that is both general and expertise-dependent. This advantage is hypothesized to depend on feedforward motor control systems that can be used during learning to transform sensory information into motor production. ■

INTRODUCTION

How do we learn sensorimotor skills such as driving, typing, speaking, or performing music? To function physically and socially, we need to move, manipulate objects, communicate, and interact with other people. Such behaviors require learning patterns of stimuli, patterns of actions, and stimulus–action relationships. Evidence from speech and music performance, skills with high cognitive and motor demands, demonstrates that, once a skill is acquired, perception and action become closely linked at the behavioral and neural levels (Brown & Palmer, 2012; Herholz & Zatorre, 2012; Saur et al., 2008; Tourville, Reilly, & Guenther, 2008; Zatorre, Chen, & Penhune, 2007; Hickok & Poeppel, 2004). Two questions remain: How do perception and action each contribute to acquiring sensorimotor skill, and what is the role of pre-existing skill or expertise? For expert speakers or musicians, hearing a new word or melody can be translated into the actions required to say or play them. For this reason, learning through listening may be equally or more effective than motor practice for experts because they can engage learned motor programs primed by sensory

information. But for novices, the sound-to-action links are not in place; thus, perceptual learning may be less effective and motor practice may be more beneficial. Based on these hypotheses, this study compares the contribution of perceptual learning to that of motor learning in sensorimotor skill acquisition for experts and novices. We tested whether trained pianists learned to play novel melodies on a piano more efficiently when learning by listening compared with performing, and we tested whether nonmusicians learned to play piano melodies more efficiently when learning by performing compared with listening.

Several theoretical perspectives point to the possibility that skilled performers may learn more efficiently via perception than novices because they can reliably translate sensory information into actions. At the cognitive level, the ability to map movements to their sensory outcomes and vice versa may occur through shared representations of action commands and their consequences (theory of event coding; Hommel, Müsseler, Aschersleben, & Prinz, 2001). Skilled performers may also be sensitive to the action relevance of sensory stimuli, even when they are not performing. For instance, when listening to our native language, we may perceive features of the actions that produce those sounds (perception-for-action-control theory; Schwartz, Basirat, Ménard, & Sato, 2012). Finally, skilled performers may be able to generalize their knowledge of

Concordia University, Montreal, Canada

*Rachel M. Brown is now at the Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, Netherlands.

stimulus–action relationships to novel stimuli through motor schemas or abstractions of stimulus–action relationships (schema theory; Schmidt, 1975). Based on the idea that skilled performers have acquired generalizable sensorimotor associations, we hypothesize that trained pianists will be able to learn novel melodies efficiently by listening alone, whereas novices may not.

At the neural level, models for speech perception and production propose specific brain networks that govern auditory–motor mapping and the control of movements based on perceived or intended auditory outcomes. Connections between superior temporal gyrus (STG) and premotor cortex (PMC) are thought to play a critical role in transforming speech sounds to motor coordinates for articulatory gestures (dual-route model; Hickok & Poeppel, 2004, 2007). The connection from PMC to primary motor cortex (M1) is thought to enable feedforward control of speech articulation or motor control without auditory feedback once speech production is learned (DIVA model; Tourville & Guenther, 2011; Guenther, 2006). Similar proposals have been made in the context of expert music performance (Zatorre et al., 2007). This type of control depends on having acquired motor programs that can be fully executed once triggered or signaled by the intended sensory outcomes. This conception fits well with feedforward models of motor control based on inverse models that simulate action commands based on the sensory outcomes of those commands (Wolpert & Kawato, 1998). Once a feedforward control system is acquired, it may allow perception to directly inform or enable action. This type of control system has been proposed to enable the fluent actions of professional musicians, which are executed too fast for auditory feedback to be useful in controlling movement (Verwey, 1999; Newell & Rosenbloom, 1981; Keele, 1968; Lashley, 1951). Therefore, in trained musicians, a premotor and primary motor cortical feedforward system would allow listening alone to access existing motor commands (Tourville & Guenther, 2011; Zatorre et al., 2007). In contrast, novices have yet to establish sensory–motor associations that enable feedforward control. Do musicians learn more efficiently by listening than by performing, and do novices learn more efficiently by executing movement than by listening?

Behavioral evidence suggests that there may be a difference between how skilled and unskilled performers weight sensory relative to motor information when learning. Skilled performers demonstrate a greater capacity to generalize acquired sensorimotor associations, but they seem to generalize the perceptual components of skill to a greater extent than the motor components. Trained pianists, for instance, demonstrate the ability to transform musical sounds to musical movements, even without awareness. Hearing piano tone patterns primed the corresponding movements for pianists, but not nonmusicians (Repp & Knoblich, 2007). Pianists also performed melodies more accurately from memory after having learned with auditory compared with no auditory

feedback, regardless of whether that feedback was present at recall (Finney & Palmer, 2003). These studies suggest that skilled performers are able to learn to perform by listening to the outcomes of actions or to outcomes that are strongly associated to actions. In addition, evidence suggests that skilled performers generalize the sensory components of acquired skill more effectively than the motor components. Adult pianists were able to transfer melody learning across different effector sequences, but they did not transfer learned movements to new melodies (Palmer & Meyer, 2000). In contrast, children with more piano experience transferred melody learning to new auditory and motor sequences, whereas children with less experience did not demonstrate either auditory or motor transfer (Palmer & Meyer, 2000). The authors argued that skilled performers “weight” perceptual components of skill more heavily than motor components when learning (p. 67). In accordance with this idea, skilled pianists performed music more accurately following listening to that music compared with performing that music without auditory feedback (Brown & Palmer, 2013). Based on this evidence, skilled performers may benefit more from perceptual learning whereas novices benefit more from motor learning in sensorimotor skill acquisition.

At the neural level, when transforming sounds to movements, musicians engage an auditory–premotor network that is similar to the one proposed to govern feedforward motor control in speech (Tourville & Guenther, 2011; Hickok & Poeppel, 2007). Musicians reliably engaged STG and PMC when listening to or imagining musical sounds, regardless of whether those sounds were novel or familiar (Brown et al., 2013; Chen, Penhune, & Zatorre, 2008a; Baumann et al., 2007; Bangert et al., 2006; D’Ausilio, Altenmüller, Olivetti Belardinelli, & Lotze, 2006), which suggests that these regions support generalizable auditory-to-motor translations. These regions may form crucial nodes within feedforward networks for skilled music performance (Brown, Zatorre, & Penhune, 2015; Zatorre et al., 2007). Nonmusicians only engage motor regions of the brain when listening to or imagining music that they had learned to play (Chen, Rae, & Watkins, 2012; Engel et al., 2012; Lahav, Saltzman, & Schlaug, 2007). Nonmusicians were also faster to learn to perform music by imitating silent videos of finger movements on a piano (visual–motor learning) compared with imitating the audio presentation of the melody (auditory–motor learning), further suggesting that nonmusicians may benefit less from auditory than from motor information when learning (Engel et al., 2012). Nonmusicians nonetheless showed greater auditory-to-visual transfer of learning than visual-to-auditory when recognizing music, suggesting that auditory learning may benefit nonmusicians’ retrieval for music (Engel et al., 2012). Taken together, this evidence suggests that, for novices, listening cannot yet guide movement as effectively as for experts.

In contrast to the hypothesis that perceptual learning may be of most benefit to trained musicians, other evidence suggests that there could be a general advantage to learning by perceiving. According to a computational neural network model of procedural learning, supported by primate data, visual–motor sequences are first encoded in visual space as a series of target positions and later in motor space as a series of joint angles (Nakahara, Doya, & Hikosaka, 2001). The authors hypothesize that perceptual learning enables efficient sequence acquisition for two reasons. One is that perceptual coordinates are more transferable than motor coordinates, because it is easier to derive the movement from a target than to derive the target from a movement (Nakahara et al., 2001). In agreement with this idea, performance on visuomotor sequences following learning can be transferred to new effectors but less easily to new perceptual sequences (Kirsch & Hoffmann, 2010; Savion-Lemieux & Penhune, 2010; Anguera, Russell, Noll, & Seidler, 2007; Japikse, Negash, Howard, & Howard, 2003; Willingham, Wells, Farrell, & Stemwedel, 2000). It is further hypothesized that the joint angles of movement require more degrees of freedom to specify than the targets of movement (Nakahara et al., 2001). Likewise, planning movements is hypothesized to require extracting only the features of a given sensory outcome that are most relevant for action (Hommel et al., 2001). Thus, a constraint on the relevant target features or degrees of freedom may enable more efficient encoding and/or more efficient translation to movement. Such a constraint may be most helpful during early acquisition of a new sensorimotor skill, when movements might be more irregular (Newell, Broderick, Deutsch, & Slifkin, 2003; Mitra, Amazeen, & Turvey, 1998). Based on these ideas, the relative advantage of perceptual versus motor learning may depend not only on skill-dependent sensorimotor associations but also on properties of perceptual information that might facilitate encoding or sensorimotor transformation. In this case, perceptual learning may be advantageous compared with motor learning for both skilled and nonskilled performers.

This study compared the contribution of perceptual learning to that of motor learning in sensorimotor skill acquisition for experts and novices. Using fMRI, we tested the degree to which pianists and nonmusicians improved their performance and engaged motor brain regions while learning to play novel piano melodies by listening compared with performing. We expected pianists to show greater improvements and response in motor regions (e.g., PMC) when listening compared with performing, and we expected nonmusicians to show greater improvements and response in motor regions when performing compared with listening.

METHODS

Participants

Thirty-two right-handed, healthy adults recruited from the McGill University community participated in the

study. All participants reported normal hearing, normal or corrected-to-normal vision, and no absolute (perfect) pitch. Half of the participants ($n = 16$) were pianists, with an average of 10.9 years of formal training (mean age = 23.6 years, 8 women), and the other half ($n = 16$) were nonmusicians, with an average of 0.4 years of formal training on any instrument or voice (mean age = 24.2 years, 11 women). The study was approved by the Montreal Neurological Research Ethics Review Board, and all participants gave written informed consent.

Equipment

During fMRI scanning, participants performed on a custom-built MRI-compatible electronic piano keyboard (Hollinger, 2008; Hollinger, Steele, Penhune, Zatorre, & Wanderley, 2007; Figure 2B). The keyboard interface comprised 24 weighted plastic piano keys from C₄ (262 Hz) to B₅ (988 Hz), though only the keys up to A₅ (932 Hz) were used in the experiment. The keyboard was attached to an adjustable plastic frame that fastened to the scanning bed. The keyboard was free of ferromagnetic parts and was connected to electronic components outside the scanner environment via fiber-optic cables. Key presses were acquired using fiber-optic sensors, which are immune to the scanner's electromagnetic interference and movable mirrors attached to each key. Sensors were composed of emitter–receiver pairs of optical fibers and were connected to a set of custom optoelectronic controller boards where light reflected by depressed keys was converted into electronic signals, analyzed, and converted to MIDI format and sent over USB to a laptop PC running Linux. Custom Python scripts were used to generate auditory feedback from keypress triggers by routing the MIDI signals to an external soundcard set to a piano timbre. Thus, participants heard low-latency, real-time auditory pitch feedback in the scanner when performing on the piano. Custom scripts also controlled the presentation of all visual and auditory stimuli generated through the external soundcard and detected trigger signals from the MR sequence. All trials were initiated by MR triggers. All sounds were presented to participants binaurally through MR-compatible Etymotic insert earphones. Sounds were adjusted to a comfortable level for each participant. All visual stimuli were presented to participants through a mirror that reflected images from a projector screen.

Stimuli

A total of 12 novel melodies were used in this study, as well as an additional practice melody. All melodies were 12 notes in length and were composed of five unique pitches, so that they could be performed using one (right) hand in a fixed position with one finger per piano key. All melodies were in 4/4 meter (one beat equaled a quarter note, with four beats per measure) and spanned 2.25 metrical measures. All melodies were presented in a

piano timbre. During scanning, each melody was preceded by four metronome beats, consisting of four 10-msec clicks presented in a drum timbre (Figure 2A). Visual cues were presented in all learning conditions in this study to enable direct comparison between auditory and motor learning. Pianists and nonmusicians learned different sets of melodies, at different speeds, and with different visual cues. Melodies were designed to be simple ($n = 4$), intermediate ($n = 4$), or complex ($n = 4$), based on the complexity of their rhythms, tonality, and contours. More complex melodies tended to have a greater number of rises and falls in the melodic line, a quantity that impacts the ability to initially encode melodies (Dowling & Fujitani, 1971). Nonmusicians learned the four simple and the four intermediate melodies. Pianists learned the four intermediate and the four complex melodies (see Figure A1 for notated versions of all melodies). Melodies were presented to pianists with a beat interonset interval (IOI) of 500 msec and to nonmusicians with a beat IOI of 800 msec (Figure 2A). Pianists were presented with standard musical notation while learning melodies, whereas nonmusicians were presented with spatial cues denoting a melody's sequence of keypresses (Figure 1). The spatial cues consisted of a horizontal array of five white squares outlined in black. Each square represented a note in the melody, and the finger of the right hand required to play that note. For

every note in the melody, one square would fill with red for the duration of that note (Figure 1).

These between-group differences in stimuli, speed, and cues were motivated by pilot testing to ensure that the learning tasks were not too difficult for nonmusicians or too easy for pianists. In particular, the use of perceptually different visual cues for each group aimed to maximize the ease with which participants could interpret the cues, given their differences in experience. These features may have engaged visual or other systems differently, and these features could have interacted with auditory and motor components of learning. However, visual cues were selected for each group according to how clearly those cues would signal how and when participants were to move their fingers. For instance, for nonmusicians the cues directly translated movement coordinates in time and space into dynamic visual coordinates (temporal pattern of horizontally arrayed finger movement). These cues were therefore expected to be the most intuitive way to convey rhythm and pitch patterns, precluding the need for training on a static notation system. Moreover, static notation must ultimately be translated into a dynamic pattern. Performance improvements for both groups reported below suggest that the visual cues were understandable for each group. Because of the presence of the visual cues, comparisons between auditory and motor conditions can be interpreted as additional influences of auditory or motor information on performance over and above visual information.

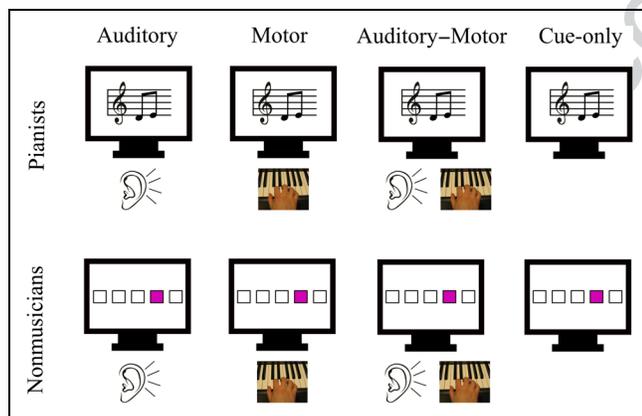


Figure 1. Learning conditions. This figure illustrates the learning conditions for pianists (top) and nonmusicians (bottom). Visual cues corresponding to the melody were presented on each learning trial in all of the learning conditions. For pianists, visual cues consisted of the melody notated in standard musical notation. For nonmusicians visual cues consisted of a horizontal array of five squares, each of which corresponded to one of the five pitches in the melody. Squares would fill with red one at a time according to the order of pitches in the melody, and each square would stay red for the duration of the corresponding pitch. The auditory condition involved viewing the visual cues while listening to the melody, but without performing the melody. The motor condition involved viewing the visual cues while performing the melody, but without listening to the melody. The auditory-motor condition involved viewing the visual cues while performing the melody and hearing auditory feedback. The cue-only condition involved viewing the visual cues without performing or listening to the melody.

Task Design and Conditions

We used a within-subject design. Each participant learned a total of eight melodies in each of four learning conditions (Figure 1), with two melodies per condition, during fMRI scanning. The cue-only condition involved viewing the visual cues corresponding to the melody (notation for pianists and spatial cues for nonmusicians) without performing or listening to the melody. The auditory condition involved viewing the visual cues while listening to the melody, but without performing the melody. The motor condition involved viewing the visual cues while performing the melody, but without listening to the melody. The auditory-motor condition involved viewing the visual cues while performing the melody and hearing auditory feedback. Both the auditory-motor and cue-only conditions served as comparisons for the two conditions of interest: auditory and motor. The auditory-motor condition was considered the standard baseline because it represents the most common way people would learn to play a melody, it provides all the required sensory and motor information, and it most closely matched test conditions at recall. The cue-only condition was considered the low-level baseline that accounted for the presence of the visual cues but provided the least information for learning. The assignment of stimuli to learning conditions and the order in which the stimuli

and conditions occurred within the experiment were counterbalanced across participants. Participants always experienced one block of each condition before the cycle of the four conditions repeated. For each melody, there were 20 learning trials involving one of the four conditions above, interleaved with a total of five recall trials, with one recall trial occurring after every four learning trials. Recall trials always involved performing the melody from memory once while hearing auditory feedback from the key presses and with no visual cues present.

Experiment trials were blocked by melody and learning condition, resulting in a total of eight experiment blocks. Each block consisted of 20 learning trials, five recall trials, as well as instruction and silence trials. Each block began with a hand position instruction trial, during which participants viewed a diagram showing where to place their hand on the piano for the duration of that block. A learning condition instruction trial followed, during which participants viewed written instructions telling them whether they were to listen and/or perform the melody during learning trials. Each block ended with two silence trials, during which participants viewed a fixation cross. The experiment was divided into two runs (four blocks each) for pianists and three runs (two runs of three blocks plus a third run of two blocks) for nonmusicians.

Procedure

Prescan

Before scanning, participants were screened, and they practiced a short version of the fMRI task with the practice melody, using the same keyboard and setup used during scanning. Before practice, participants were trained to place their right hand correctly on the keyboard for each melody, as indicated by diagrams presented on a computer screen, without looking at their hands on the piano. Participants learned four hand positions, one for each musical key in which the melodies were presented (two melodies per key). Velcro markers placed in the thumb positions served as tactile cues during practice and the experiment. Participants could place the right thumb on the marker and the other four fingers on adjacent white keys.

Scan

The keyboard was secured to the scanning bed at a comfortable arm's length for the participant. Padding was placed around the participant's right arm and head to minimize movement. Participants completed an anatomical scan, followed by the functional scanning runs. Before functional scanning, participants were reminded of the hand position for each musical key and executed each of them once to ensure accuracy. All keystrokes and keystroke onsets and offsets during learning and recall trials were recorded online.

fMRI Acquisition

Scanning was performed on a 3-T Siemens Sonata Imager with a 32-channel head coil. A high-resolution T1-weighted structural image was first acquired for each participant (MPRAGE sequence, voxel size = $1 \times 1 \times 1 \text{ mm}^3$, matrix size = 256×192 , echo time = 2.98 msec, repetition time [TR] = 2300 msec, flip angle = 9°). Two (pianists) or three (nonmusicians) functional runs were then acquired using a T2*-weighted gradient-echo EPI sequence. For pianists, each functional run contained 119 volumes. For nonmusicians, the first and second runs each contained 90 volumes, and the third run contained 59 volumes. Each volume contained 40 whole-head interleaved slices oriented perpendicular to the Sylvian fissure (echo time = 30 msec, voxel size = $3.5 \times 3.5 \times 3.5 \text{ mm}^3$, matrix size = 64×64). A sparse-sampling paradigm was used, which minimizes the influence of scanner noise on task-related BOLD response and takes advantage of the delay in the hemodynamic response following a stimulus or event (Gaab, Gabrieli, & Glover, 2007; Belin, Zatorre, Hoge, Evans, & Pike, 1999; Glover, 1999). Volumes took 2400 msec to acquire, and they were acquired every 9500 msec (TR = 9500 msec) for pianists and every 13,600 msec (TR = 13600 msec) for nonmusicians. This TR difference between groups was put in place to present stimuli more slowly to nonmusicians, as motivated by pilot testing (discussed above). Because the absolute time delay between the end of stimulus presentation and the beginning of slice acquisition was similar for both TR times (0.8 sec for nonmusicians and 0.5 sec for pianists), we expected both acquisition sequences to be sensitive to hemodynamic response following stimulus presentation. Stimulus presentation or performance occurred in the 4500 or 7200 msec between metronome beats and scan acquisitions (Figure 2A). Field maps were also acquired and used in analysis to correct functional data for image distortions.

Behavioral Analysis

Pitch and rhythm accuracy were calculated for each recall trial for each participant. Pitch accuracy was calculated as the percentage of pitches performed in the correct order. Any pitches omitted, substituted, or played in the incorrect order were counted as errors. Rhythm accuracy was calculated as the percentage of IOI durations performed in the correct order. To account for tempo drifts, correct IOIs were defined as those which fell within a maximum of a 30% difference from an expected IOI for that interval (Drake & Palmer, 2000). The expected IOI was based on the nearest previous correct IOI that was performed and the rhythmic category (quarter note, eighth note, etc.) of that IOI. For instance, if a participant correctly performed a quarter note interval of 450 msec, a subsequent eighth note IOI between 150.75 and 299.25 msec would be correct. Pitch and rhythm accuracy improvements across recall trials were assessed using the area under the curve

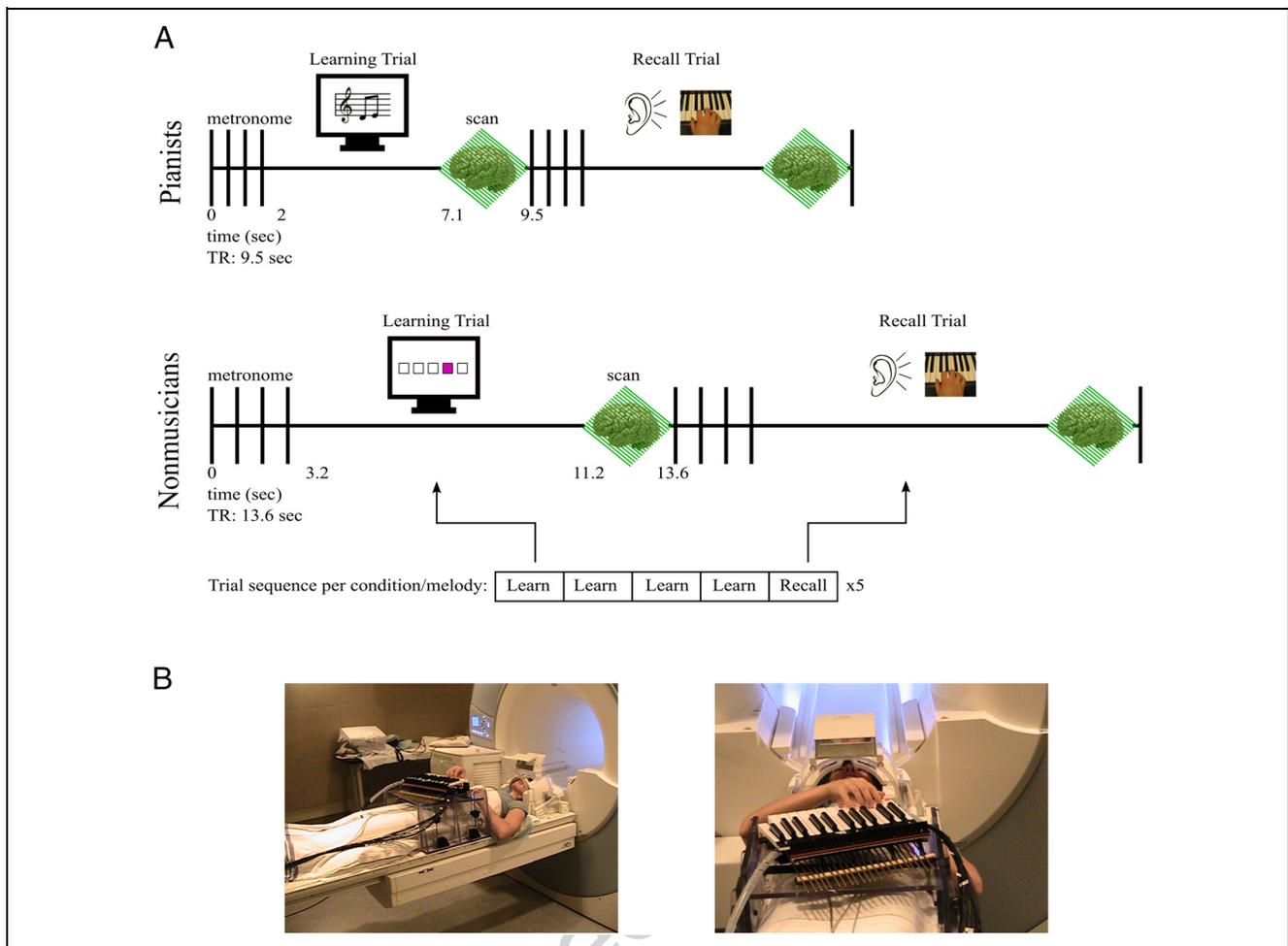


Figure 2. Scanning paradigm and MR-compatible piano. (A) These illustrations show the scanning paradigm. Each learning and recall trial began with four metronome beats, with a beat IOI of 500 msec for pianists and 800 msec for nonmusicians. Participants were instructed to begin playing after the fourth metronome beat on recall trials. Each trial lasted 9.5 sec for pianists (2 sec for the metronome beats, 4.5 sec for melody presentation or recall, a 0.6-sec silence buffer, and 2.4 sec for the scan acquisition) and 13.6 sec for nonmusicians (3.2 sec for the metronome beats, 7.2 sec for melody presentation or recall, a 0.8-sec silence buffer, and 2.4 sec for the scan acquisition). (B) These pictures show the MR-compatible piano keyboard in the scanning environment.

(AUC), which describes repeated measurements over time (Fekedulegn et al., 2007; Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). AUC retains information from all measurements at each time point over learning and makes no assumptions about the shape of the learning curve (Pruessner et al., 2003). AUC was computed for pitch and rhythm accuracy for each participant per stimulus and condition. AUC was computed using the trapezoid formula (Pruessner et al., 2003) or the sum of areas defined by subsequent measurements (e.g., Recall Trial 1 and Recall Trial 2) and accuracy changes between subsequent measurements (e.g., 50%–60% accuracy) relative to a baseline accuracy of zero (e.g., 50%–0% and 60%–0%). We chose a zero baseline because the first accuracy measurement at Recall Trial 1 occurred after a set of learning trials. AUC values represent units of accuracy (% correct) multiplied by the recall trial number. Possible values range from zero (0% accuracy \times 5 trials = 0), signifying zero accuracy on all trials, to 500 (100%

accuracy \times 5 trials = 500), signifying perfect accuracy on every trial. For instance, a value of 350 signifies that 70% of the maximal accuracy over trials was achieved ($350/500 \times 100 = 70$).

Linear mixed-effects (LME) models assessed changes in pitch and rhythm accuracy across recall trials in each learning condition and AUC for pitch and rhythm accuracy across learning conditions. One set of LME models compared pitch and rhythm accuracy by learning condition and recall trial and included two fixed effects (four learning conditions and five recall trials) and two random effects (subject and stimulus). Another set of LME models compared AUCs for pitch and rhythm accuracy across learning conditions. These included one fixed effect (four learning conditions) and two random effects (subject and stimulus). Planned comparisons assessed pairwise differences between recall trials and learning conditions. In addition, two additional sets of LME models were run for each analysis above to address the

potential effects of the order in which conditions and stimuli occurred in the experiment. To account for variance due to order effects, one set of LME models included the order of conditions/stimuli as a random factor. To explicitly examine order effects, a second set of LME models included order of conditions/stimuli as a fixed factor. Unless otherwise indicated, results remained unchanged with order included as a random factor, and no fixed effects (either main effects or interaction effects) of order were found. Statistical analyses were run in R (Version 3.2.2).

fMRI Analysis

Functional MRI data were analyzed using FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl; Smith et al., 2004). Functional images were motion-corrected using MCFLIRT (Motion Correction FMRIB Linear Registration Tool; Jenkinson, Bannister, Brady, & Smith, 2002), unwarped using a field map, and spatially smoothed using a Gaussian kernel of 8 mm FWHM. High-pass filters of 256.5 sec for pianists and 367.2 sec for nonmusicians were used to remove low-frequency drift. Nonbrain tissue was removed from functional and anatomical scans using BET (Brain Extraction Tool; Smith, 2002). Functional images were registered to their respective structural images using FLIRT (FMRIB's Linear Registration Tool; Jenkinson et al., 2002; Jenkinson & Smith, 2001) with 3 degrees of freedom, and structural images were registered to MNI-152 standard space using linear registration with 12 degrees of freedom. The volumes of each functional run corresponding to instruction trials were discarded from analyses.

The general linear model (GLM) was used to compute statistical maps of activity corresponding to contrasts between each learning condition compared with silence and between learning conditions, averaged across learning and recall trials separately. Learning and recall trials in each learning condition were modeled separately by assigning all learning or recall trials in each condition a coefficient of 1 and all other condition and silence trials a coefficient of 0, resulting in eight regressors. Thus, z -statistic maps for each of the eight parameter estimates represented increased BOLD response compared with silence during learning or recall trials for one of the four learning conditions. In addition, BOLD increases and decreases over learning and recall trials were examined using two models. One model examined increases and decreases across each block of the task in a particular condition (1 block = 4 Learning trials or 1 recall trial). Each group of four learning trials that preceded a recall trial was assigned to a coefficient of -2 , -1 , 0 , 1 , or 2 for Blocks 1 through 5, respectively (model of increase over block) or a coefficient of 2 , 1 , 0 , -1 , -2 for Blocks 1 through 5, respectively (model of decrease over block). The second model examined increases and decreases

across early and late blocks of a condition (early = Blocks 1 and 2, late = Blocks 4 and 5). Learning and recall trials were assigned to coefficients -1 or 1 for early versus late blocks (increases) or 1 and -1 for early versus late blocks (decreases) in each condition. Each of the above analyses were first performed for each run per participant and then averaged across runs per participant. Group averages were obtained by submitting each participant's activation map into a Stage 1 and 2 group analysis in FLAME (FMRIB's Local Analysis of Mixed Effects; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004). Z -statistic images were thresholded using clusters determined by $Z > 2.3$ and a corrected cluster significance threshold of $p = .05$ (Worsley, 2001). Localization was determined using the Juelich histological atlas (Eickhoff et al., 2007), the Harvard-Oxford cortical and subcortical structural atlases, and the cerebellar atlas that are part of the FSL software.

Group-level differences between auditory and motor learning were examined using contrasts between pairs of conditions and compound contrasts for learning and recall trials and for pianists and nonmusicians separately. The auditory and motor conditions were contrasted with either the cue-only (Cue) or auditory-motor (AM) conditions (Auditory $>$ Cue, Motor $>$ Cue, Auditory $>$ AM, Motor $>$ AM) and then contrasted with each other, yielding four compound contrasts: (1) (Auditory $>$ Cue) $>$ (Motor $>$ Cue), (2) (Auditory $>$ AM) $>$ (Motor $>$ AM), (3) (Motor $>$ Cue) $>$ (Auditory $>$ Cue), and (4) (Motor $>$ AM) $>$ (Auditory $>$ AM). Follow-up ROI analyses compared BOLD response to performance accuracy in the auditory condition. Separate ROI masks for pianists and nonmusicians comprised a sphere with a 7-mm radius and were centered on peak voxels from contrasts above. For pianists, an ROI in left primary motor cortex was chosen from the compound contrast (Auditory-AM) $>$ (Motor $>$ AM) for recall trials. For both pianists and nonmusicians, ROIs in left inferior parietal cortex (IP) were chosen from the (Auditory $>$ AM) contrast for learning trials. For nonmusicians, two ROIs in left STG (one from a middle cluster and one from a posterior cluster) were chosen from the (Auditory $>$ Cue) contrast for learning trials. Mean percent BOLD response change was extracted from each ROI for learning or recall trials in the auditory condition. Using R, BOLD response change was correlated with both pitch and rhythm accuracy during recall trials in the auditory condition. Functional connectivity was also examined in the auditory condition using psychophysiological interaction (PPI) analyses. For pianists, one seed in right STG was chosen from the contrast (Auditory $>$ Cue) $>$ (Motor $>$ Cue) for learning trials, and one seed in left primary motor cortex was chosen from the contrast (Auditory-AM) $>$ (Motor $>$ AM) for recall trials. For nonmusicians, the three ROI regions selected above were also selected as seed regions. Seeds were constructed in the same way as ROI masks. The BOLD response time series from the seed regions, and the interactions between the

seed time series and the GLM regressors for auditory learning trials and auditory recall trials were added as regressors to the GLM, along with the original regressors representing learning and recall trials in each of the four conditions. This model was tested for each run per participant and subsequently averaged across runs and across participants, separately for pianists and nonmusicians. Finally, conjunction analyses examined the spatial overlap between pianists' and nonmusicians' thresholded maps for auditory and motor conditions. Based on peak voxels from these conjunctions, BOLD response change was extracted from five ROIs in primary motor cortex, STG, and IP. Response in these ROIs was compared across pianists and nonmusicians in auditory and motor conditions using R.

RESULTS

Behavioral Results

Pianists: Pitch and Rhythm Accuracy

Pianists showed pitch and rhythm accuracy improvement across recall trials in all learning conditions, and they showed greater average pitch and rhythm accuracy across all recall trials in the auditory compared with the motor condition (Figure 3; see Figure A2 for individual pitch accuracy learning curves). An LME model for pitch accuracy revealed main effects of recall trial and learning condition (recall trial, $F(4, 598) = 38.41$, generalized eta-

squared (η^2) = 0.25, $p < .05$, Recall Trial 1 < Recall Trials 2, 3, 4, and 5, $ps < .05$, Recall Trial 2 < 5, $p = .087$; learning condition, $F(3, 598) = 14.49$, $\eta^2 = 0.08$, $p < .05$, Auditory > Motor and AM conditions, $ps < .05$, Auditory > Cue, $p = .089$). An LME model for rhythm accuracy also revealed main effects of recall trial and learning condition (recall trial, $F(4, 598.01) = 15.98$, $\eta^2 = 0.11$, $p < .05$, Recall Trial 1 < Recall Trials 3, 4, and 5, $ps < .05$; learning condition, $F(3, 598.01) = 7.79$, $\eta^2 = 0.04$, $p < .05$, Auditory > Motor condition, $p < .05$, Auditory > Cue, $p = .071$). The magnitude of pitch and rhythm accuracy improvement differed among the learning conditions (Figure 3). An LME model for pitch accuracy AUC revealed a main effect of learning condition ($F(3, 102.03) = 5.46$, $\eta^2 = 0.12$, $p < .05$, Auditory > Motor, AM conditions, $ps < .05$). An LME model for rhythm accuracy AUC also revealed a main effect of learning condition ($F(3, 102.01) = 3.92$, $\eta^2 = 0.06$, $p < .05$, Auditory > Motor condition, $p < .05$, Auditory > AM, $p = .080$, Auditory > Cue, $p = .061$). In summary, pianists showed greater pitch and rhythm accuracy improvement during auditory compared with motor learning.

Nonmusicians: Pitch and Rhythm Accuracy

Nonmusicians demonstrated improvement in pitch and rhythm accuracy across recall trials in all learning

Figure 3. Pianists' pitch and rhythm accuracy and accuracy improvement by learning condition. The top panel illustrates pitch accuracy (percent correct pitches in the correct order) by recall trial and pitch accuracy improvement (area under the learning curve or AUC) by learning condition. The bottom illustrates rhythm accuracy (percent correct durations in the correct order) by recall trial and rhythm accuracy improvement (AUC) by learning condition. Error bars represent standard error.

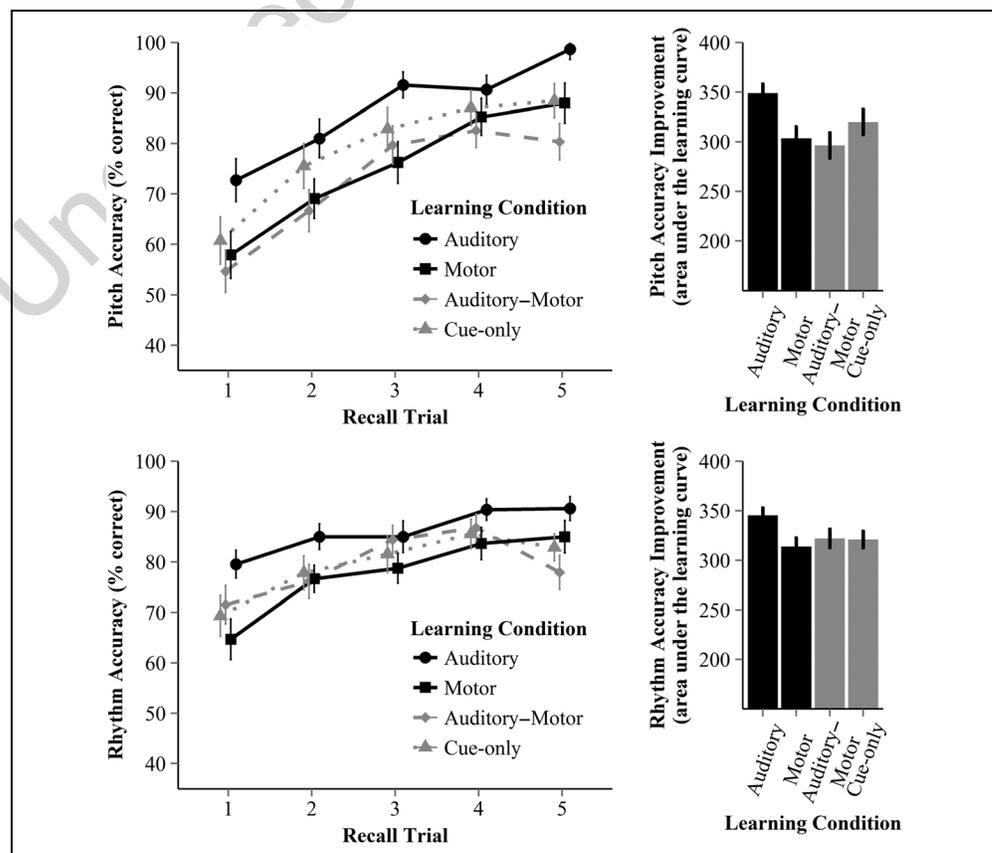
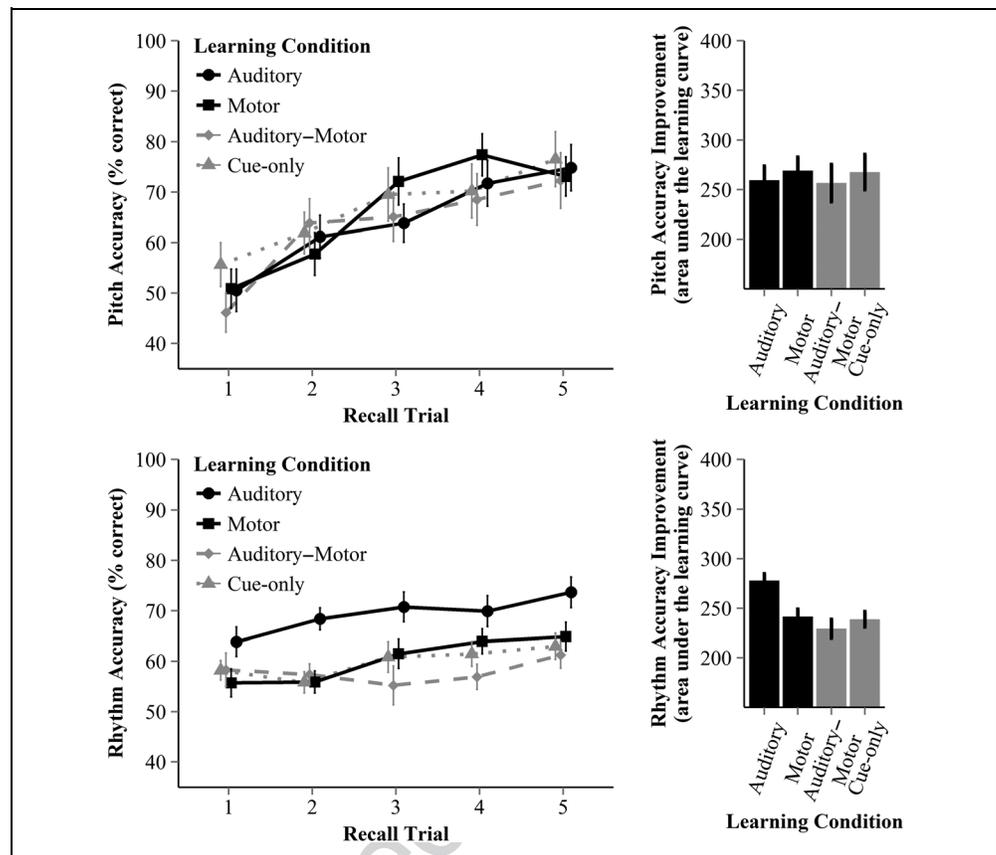


Figure 4. Nonmusicians' pitch and rhythm accuracy and accuracy improvement by learning condition. The top panel illustrates pitch accuracy (percent correct pitches in the correct order) by recall trial and pitch accuracy improvement (AUC) by learning condition. The bottom illustrates rhythm accuracy (percent correct durations in the correct order) by recall trial and rhythm accuracy improvement (AUC) by learning condition. Error bars represent standard error.



conditions, but they only demonstrated greater average rhythm accuracy across recall trials in the auditory compared with the motor condition (Figure 4; see Figure A3 for individual pitch accuracy learning curves). An LME model for pitch accuracy revealed a main effect of recall trial but not learning condition ($F(4, 598) = 24.14, \eta^2 = 0.17, p < .05$, Recall Trial 1 < Recall Trial 5, $p < .05$, Recall Trial 1 < 4, $p = .057$, Recall Trial 1 < 3, $p = .080$, Recall Trial 2 < 5, $p = .055$). An LME model for rhythm accuracy also revealed main effects of recall trial and learning condition, though pairwise contrasts did not reach significance (recall trial, $F(4, 597.99) = 4.33, \eta^2 = 0.04, p < .05$; learning condition, $F(3, 600.63) = 19.34, \eta^2 = 0.10, p < .05$). In an additional set of analyses including order of stimulus/condition as a fixed factor, order showed a main effect on rhythm accuracy ($F(7, 436.11) = 2.47, p < .05, \eta^2 = 0.03$), though no pairwise differences among orders reached significance. The magnitude of rhythm accuracy improvement (rhythm AUC) differed among the learning conditions, but not the magnitude of pitch accuracy (Figure 4). An LME model on rhythm AUC revealed a main effect of learning condition ($F(3, 103.01) = 9.10, \eta^2 = 0.16, p < .05$, Auditory > Motor, Cue and AM conditions, $ps < .05$). In summary, nonmusicians demonstrated greater rhythm accuracy improvement in the auditory learning condition compared with all other learning conditions.

Pianists and Nonmusicians: Pitch and Rhythm Accuracy

Pitch and rhythm accuracy were compared across groups only for those melodies that both groups performed (melodies of “intermediate” complexity) using mixed-effects ANOVAs on pitch and rhythm accuracy as well as AUC values. Overall, these comparisons suggest that pitch accuracy was generally more difficult to acquire in the auditory-motor condition compared with the other conditions. These results also confirm that, for both groups, auditory learning was advantageous for rhythm learning, though pianists showed greater rhythm accuracy improvements over recall trials than nonmusicians.

Main effects of group, condition, and recall trial for both pitch and rhythm accuracy demonstrate greater accuracy for pianists, greater accuracy in auditory, motor, and cue conditions relative to the auditory-motor condition, and increases in accuracy over recall trials (pitch accuracy: group [$F(1, 30) = 28.64, p < .05, \eta^2 = 0.22$, Pianists > Nonmusicians, $p < .05$], recall trial [$F(3.08, 92.43) = 33.92, p < .05, \eta^2 = 0.10$, Recall Trial 1 < 2, 3, 4, and 5, Recall Trial 2 < 3, 4, 5, $ps < .05$], learning condition [$F(2.54, 76.20) = 3.97, p < .05, \eta^2 = 0.04$, Auditory > AM, Motor > AM, $ps < .05$, Cue > AM, $p = .063$]; rhythm accuracy: group [$F(1, 30) = 62.46, p < .05, \eta^2 = 0.31$, Pianists > Nonmusicians, $p < .05$], recall trial

[$F(3.42, 102.67) = 7.00, p < .05, \eta^2 = 0.03$, Recall Trial 1 < 4, 5, Recall Trial 2 < 4, 5, $ps < .05$], learning condition [$F(2.15, 64.61), p < .05, \eta^2 = 0.05$, Auditory > Motor, Cue, AM, $ps < .05$]. An additional interaction between group and recall trial for rhythm accuracy showed greater rhythm accuracy improvement over trials for pianists compared with nonmusicians (rhythm accuracy: group by recall trial, $F(3.42, 102.67) = 4.44, p < .05, \eta^2 = 0.02$; Pianists: Recall Trial 1 < 5, $p < .05$, Nonmusicians: Recall Trial 1 < 5, ns).

Main effects of group and condition on the magnitude of pitch and rhythm accuracy improvement (AUC) showed greater accuracy improvements for pianists and greater accuracy improvements in the auditory, motor, and cue conditions compared with the auditory–motor condition (pitch AUC: group [$F(1, 30) = 28.08, p < .05$,

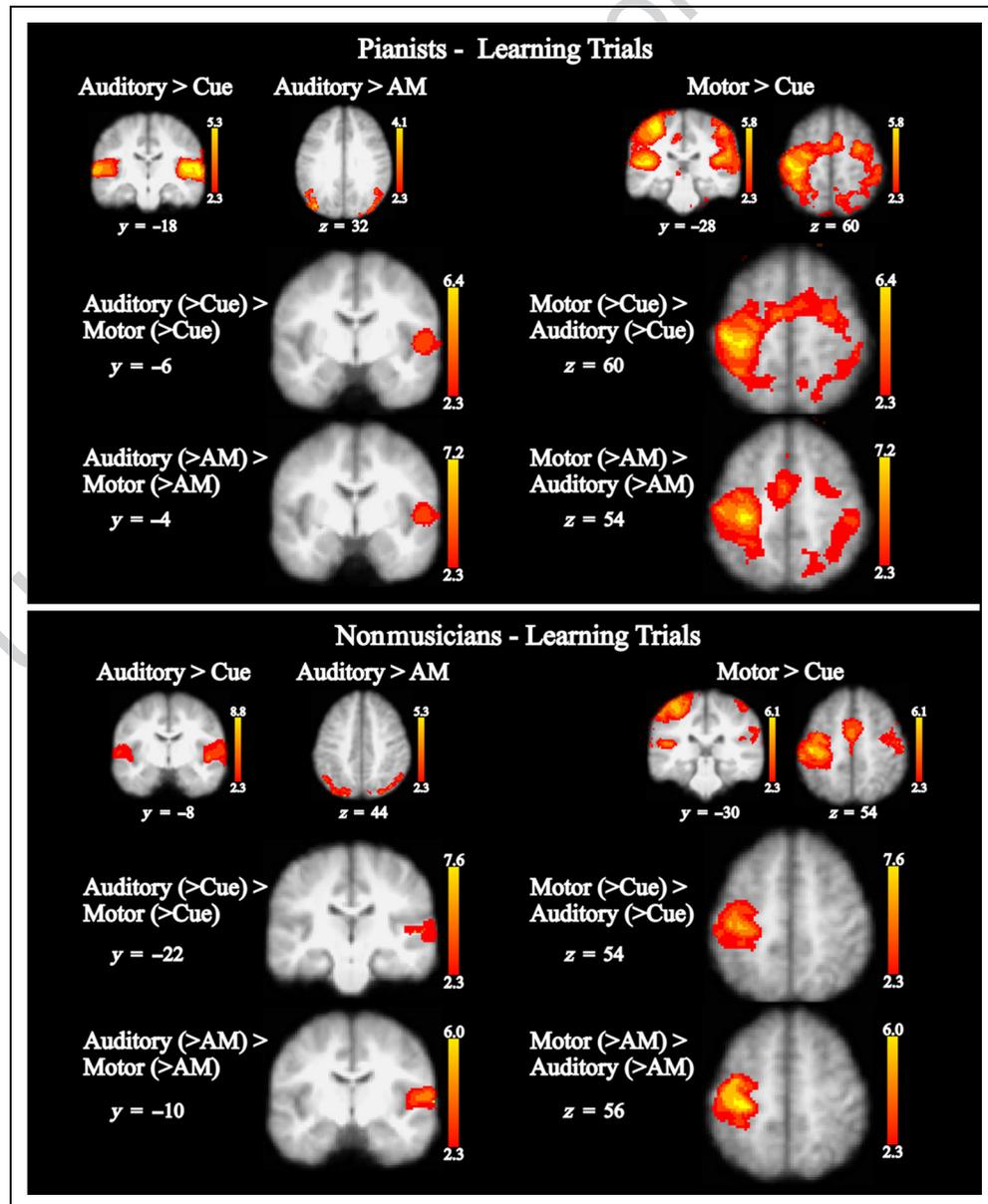
$\eta^2 = 0.30$, Pianists > Nonmusicians, $p < .05$], condition [$F(2.59, 77.81) = 3.68, p < .05, \eta^2 = 0.04$, Auditory > AM, Motor > AM, $ps < .05$, Cue > AM, $p = .073$]; rhythm AUC: group [$F(1, 30) = 68.59, p < .05, \eta^2 = 0.50$, Pianists > Nonmusicians, $p < .05$], condition [$F(2.13, 63.93) = 5.43, p < .05, \eta^2 = 0.05$, Auditory > Motor, Cue, and AM conditions, $ps < .05$]. There were no interactions between group and learning condition.

fMRI Results

Pianists: Average BOLD Response during Learning and Recall Trials

Across all learning trials for pianists, a frontal–parietal–subcortical network responded in each condition

Figure 5. Mean BOLD response during learning trials for pianists and nonmusicians. These statistical maps show regions that were engaged during auditory learning and motor learning and regions that were more engaged during auditory learning compared with motor learning and vice versa. “Cue” refers to the cue-only condition, and “AM” refers to the auditory–motor condition. For both pianists (top) and nonmusicians (bottom), right STG was more engaged during auditory compared with motor learning, and left M1 was more engaged during motor compared with auditory learning. Maps are cluster-thresholded at $z = 2.3, p < .05$.



compared with silence. This network, common to learning trials in each condition, included SMA and PMC, superior parietal cortex (SP) and IP, and the cerebellum. This network also included STG. STG responded during auditory and auditory–motor learning trials, as well as during motor and cue learning trials (left STG), although STG response was less extensive in the motor condition and much less so in the cue condition. In addition, contrasts between conditions showed mainly auditory and motor cortical activity related to auditory and motor conditions, respectively. Greater response during auditory learning trials compared with cue and auditory–motor learning trials appeared in bilateral STG (Auditory > Cue) and bilateral IP (Auditory > AM). Greater response during motor learning trials compared with cue learning trials (Motor > Cue) appeared in left primary motor (M1) cortex, bilateral primary somatosensory (S1) cortex, right secondary somatosensory cortex (S2), premotor regions (SMA and PMC), parietal regions (SP), the cerebellum, as well as STG. Finally, compound contrasts between auditory and motor learning trials further demonstrated auditory and motor cortical response associated with auditory and motor learning trials, respectively. Right STG response was greater during auditory compared with motor learning trials ((Auditory > Cue) > (Motor > Cue) and (Auditory > AM) > (Motor > AM)). Response in a number of primary motor, premotor, parietal, and subcortical regions was greater during motor compared with auditory learning trials, including left M1, bilateral S1, PMC, SMA, SP, cerebellum ((Motor > Cue) > (Auditory > Cue)), left M1, and SMA ((Motor > AM) > (Auditory > AM); Figure 5; Table 1).

Across all recall trials for pianists, a similar frontal–parietal–subcortical network responded in each learning condition compared with silence, and this network also included auditory and motor cortical regions (because recall always involved performing with auditory feedback). This network, common to recall trials in each learning condition, included SMA, PMC, M1, S1, SP, STG, and cerebellum. Contrasts between conditions only showed greater response during recall trials in the auditory condition compared with other conditions, and this response appeared in motor cortical regions. Greater response during recall trials in the auditory compared with the auditory–motor condition appeared in left M1 and S1 (Auditory > AM). Similarly, greater response during recall trials in the auditory compared with the motor condition also appeared in left M1 ((Auditory > Cue) > (Motor > Cue)) and both left M1 and S1 ((Auditory > AM) > (Motor > AM); Figure 6; Table 1).

Pianists: Decrease in BOLD Response over Learning and Recall Trials

Across all blocks and across early to late blocks of learning trials, BOLD response decreased in each condition compared with silence. Response decreased during

learning trials in frontal, premotor, parietal, and subcortical regions. These regions included ventrolateral pFC, dorsal and ventral PMC (vPMC), SMA, SP and IP, cerebellum, as well as visual cortex. In auditory learning trials specifically, response decreases appeared in STG. In addition, contrasts between conditions revealed greater decreases during auditory learning trials compared with learning trials in the cue-only and auditory–motor conditions. These greater decreases during auditory learning appeared not only in auditory cortical areas but in premotor and parietal areas as well. Greater decreases across all blocks during auditory learning trials compared with cue and auditory–motor learning trials appeared in right STG (Auditory > Cue) and in right STG, vPMC, and IP (Auditory > AM). Greater decreases from early to late blocks during auditory learning trials compared with cue and auditory–motor learning trials appeared in S2 and right STG (Auditory > Cue) and in right STG, right S1, vPMC, SP, cerebellum, and visual cortex (Auditory > AM) (Figure 7; Table 2).

BOLD response also decreased over recall trials, across all blocks, and across early to late blocks, in each condition compared with silence. Decreases appeared mainly in pre-SMA and cingulate cortex. Additional decreases from early to late blocks appeared in specific conditions compared with silence. These included parietal cortex during auditory and motor recall, temporal-occipital cortex, visual cortex, and cerebellum during auditory and cue recall, and additionally frontal operculum and PMC during auditory recall. In addition, contrasts between conditions revealed greater decreases during recall trials in the auditory condition. Greater decreases across all blocks and early to late blocks in the auditory compared with the auditory–motor condition appeared in right S1, left cerebellum, and visual cortex (Auditory > AM) (Table 2).

Pianists: BOLD–Performance Correlations

To further examine the relationship between primary motor response and performance improvements, an ROI was chosen in left M1 from the compound contrast (Auditory-AM) > (Motor > AM) for recall trials. BOLD response in left M1 for each pianist during learning and recall trials in the auditory condition (contrasted with the auditory–motor condition) was compared with each pianist’s magnitude of pitch and rhythm accuracy improvement across recall trials (AUC). A negative correlation was found between BOLD response in M1 during auditory learning trials and pitch accuracy improvement over recall trials ($r = -.50, p = .0505$), whereby pianists who showed smaller average BOLD response in M1 during auditory learning showed greater pitch accuracy improvements (Figure 6).

In addition, the relationship between IP response and performance was examined. IP response was of interest because it is a multimodal region that was engaged more

Table 1. Pianists: Mean BOLD Response during Learning and Recall Trials

<i>Contrast</i>	<i>Region</i>	<i>Learning Trials</i>		<i>Recall Trials</i>		
		<i>Z</i>	<i>(x y z)</i>	<i>Z</i>	<i>(x y z)</i>	
Auditory > Cue	Right STG	5.36	(56 -24 10)			
	Left STG	4.97	(-62 -12 8)			
Motor > Cue	Left M1	5.69	(-32 -28 60)			
	Left S1	5.46	(-48 -20 50)			
	Left SP	4.11	(-28 -58 66)			
	Left PMC	4.99	(-24 -14 64)			
	Left STG	5.50	(-46 -28 12)			
	Left cerebellum	5.35	(-20 -54 -22)			
	SMA	4.25	(8 -4 66)			
	Right S1	4.67	(48 -28 54)			
	Right SP	3.51	(20 -58 64)			
	Right PMC	4.63	(28 -8 64)			
	Right S2	5.26	(50 -22 16)			
	Right VC	5.67	(20 -56 -10)			
	Right cerebellum	5.66	(10 -64 -40)			
	Auditory > AM	Left IP	4.06	(-40 -78 32)		
Right IP		3.63	(44 -76 36)			
Left M1				3.96	(-40 -20 56)	
Left S1				4.61	(-50 -18 50)	
Auditory (>Cue) > Motor (>Cue)	Right STG	4.57	(50 -6 10)			
	Left M1			4.28	(-38 -28 64)	
Motor (>Cue) > Auditory (>Cue)	Left M1	6.46	(-32 -28 60)			
	Left S1	5.94	(-46 -20 50)			
	Left SP	4.34	(-34 -48 60)			
	Left PMC	4.59	(-26 -14 60)			
	SMA	4.3	(-2 -6 52)			
	Right cerebellum	5.7	(24 -42 -24)			
	Right S1	4.42	(48 -28 54)			
	Right SP	3.42	(24 -64 48)			
	Right PMC	4.30	(28 -10 62)			
	Auditory (>AM) > Motor (>AM)	Right STG	3.79	(54 -4 4)		
		Left M1			4.17	(-38 -26 60)
Left S1				4.65	(-50 -14 48)	
Motor (>AM) > Auditory (>AM)	Left M1	7.12	(-32 -28 54)			
	SMA	4.3	(-2 -6 52)			

Table 1. (continued)

Contrast	Region	Learning Trials		Recall Trials	
		Z	(x y z)	Z	(x y z)
Auditory, Seed: M1 (-38, -28, 64)	Left STP	4.40	(-56 -44 22)		
	Left STS	3.69	(-54 -32 0)		
	Left vPMC	3.76	(-54 -2 10)		
	Right vPMC	3.93	(52 6 6)		

Coordinates for peak activations are in MNI-152 space, and peak z values are significant at $p < .05$, corrected. Cue = cue-only condition; AM = auditory-motor condition; M1 = primary motor cortex; S1 = primary somatosensory cortex; S2 = secondary somatosensory cortex; VC = visual cortex.

during listening compared with listening plus performance (the Auditory > AM contrast for learning trials). Response in this region during learning trials (contrasted with the auditory-motor condition) also correlated negatively with pitch accuracy improvement across recall trials, albeit more weakly ($r = -.47$, $p = .067$).

Pianists: Functional Connectivity

To further probe how the auditory condition influenced connectivity among auditory and motor regions, both right STG and left M1 were chosen as seed regions for a functional connectivity analysis. Right STG was chosen from peak activations from compound contrasts showing

Figure 6. BOLD response associated with auditory learning for pianists and nonmusicians. The top left panel illustrates the marginally significant correlation between response in left M1 (primary motor cortex) during auditory learning and pitch accuracy improvement (area under the learning curve) in the auditory condition for pianists. This panel also illustrates the auditory (STG/STS) and premotor (vPMC) regions that showed functional connectivity with left M1 (seed region centered on voxel -38, -28, 64) during auditory learning trials for pianists. The top right panel illustrates the greater response in left M1 during recall trials in the auditory compared with the motor condition for pianists. The bottom panel illustrates the greater response in IP in the auditory compared with the auditory-motor condition for nonmusicians. This panel also illustrates the correlation between IP response during auditory learning trials and pitch accuracy improvement in the auditory condition for nonmusicians. “Cue” refers to the cue-only condition, and “AM” refers to the auditory-motor condition. Maps were cluster-thresholded at $z = 2.3$, $p < .05$.

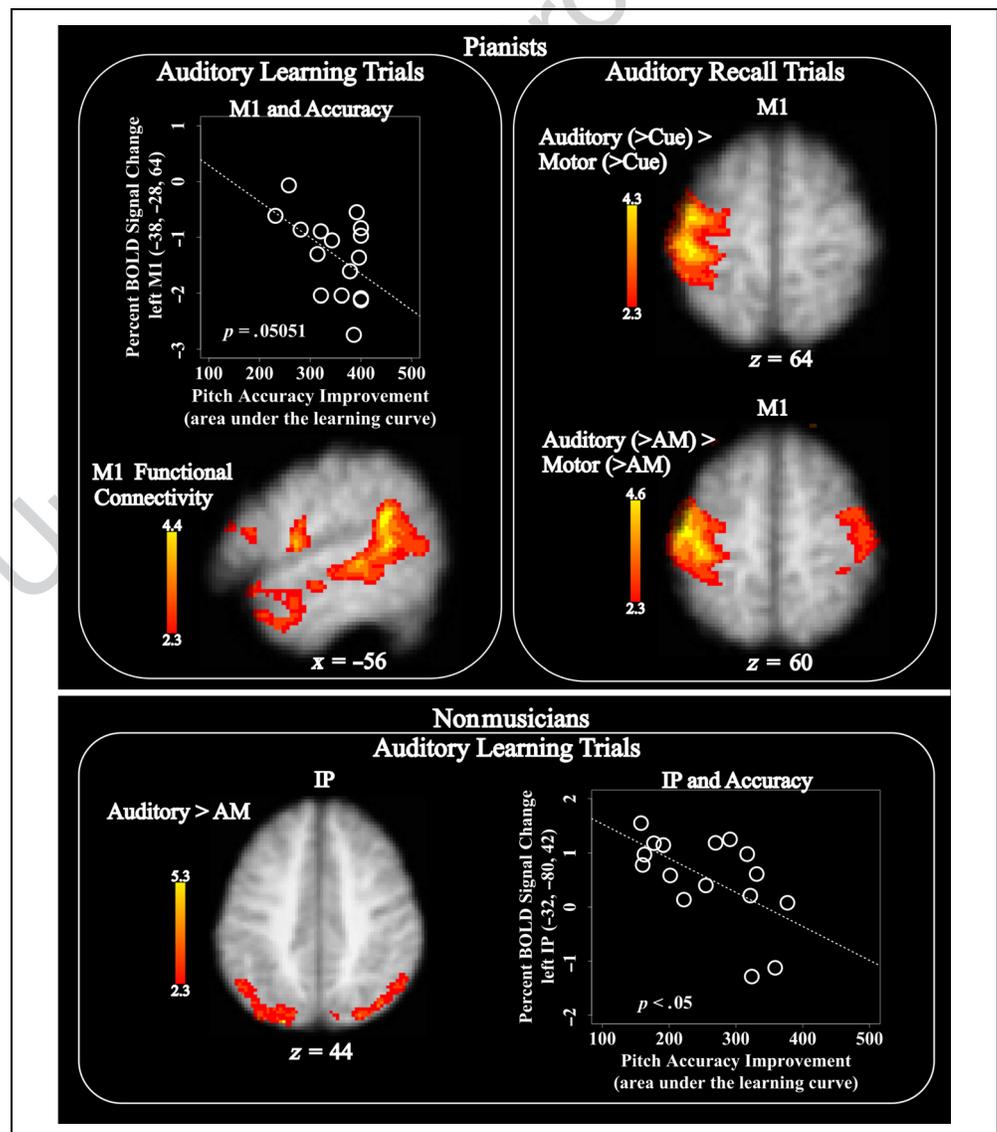
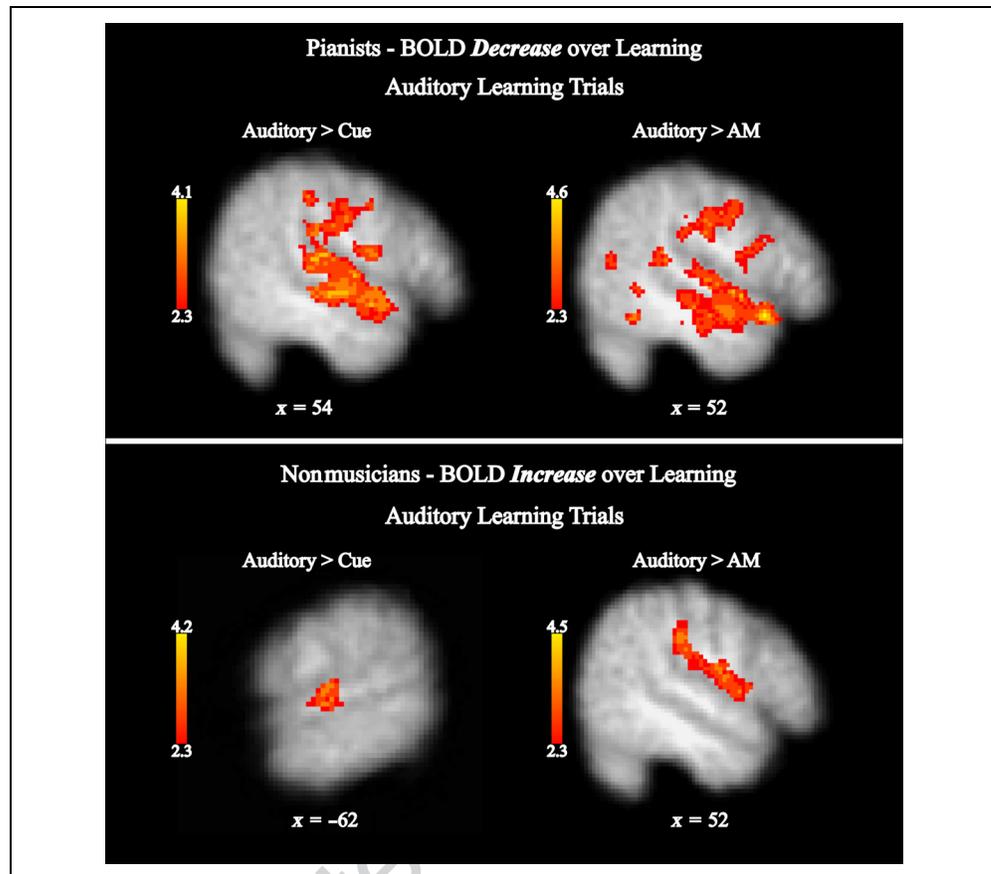


Figure 7. BOLD response decreases and increases associated with auditory learning for pianists and nonmusicians, respectively. Decreases and increases shown are across early to late blocks of learning trials. The top panel illustrates greater BOLD response decreases for pianists from early to late learning blocks in somatosensory (S2) and auditory (right STG) regions in the auditory compared with the cue-only condition, as well as greater decreases from early to late learning blocks in somatosensory (S1), premotor (vPMC), parietal (SP), and auditory regions (right STG) in the auditory compared with the auditory–motor condition. The bottom panel illustrates greater BOLD response increases for nonmusicians from early to late learning blocks in auditory regions (left STG) in the auditory compared with the cue-only condition and greater increases in somatosensory regions (S2) in the auditory compared with the auditory–motor condition. “Cue” refers to the cue-only condition, and “AM” refers to the auditory–motor condition. Maps were cluster-thresholded at $z = 2.3, p < .05$.



greater STG response for learning trials in the auditory versus motor condition, and M1 was chosen from peak activations from compound contrasts showing greater M1 response during recall trials in the auditory versus motor condition. Correlations with these seed regions in the auditory condition during learning and recall trials across the whole brain were examined. No regions showed above-threshold correlations with the RSTG seed region. Correlations with M1 response in the auditory condition for learning trials revealed peak clusters in left superior temporal-parietal junction (area STP), left STS, bilateral vPMC, and right occipital cortex (Figure 6; Table 1).

Nonmusicians: Average BOLD Response during Learning and Recall Trials

Across all learning trials for nonmusicians, a frontal-parietal network responded in each condition compared with silence. This network, common to learning trials in each condition, included SMA, PMC, S1, and SP. This network also included STG. STG responded during auditory

and auditory–motor learning trials and during motor and cue learning trials; in the latter conditions, STG response was less extensive and more posterior than in the conditions with auditory stimulation. In addition, contrasts between conditions showed mainly auditory and motor cortical activity related to auditory and motor conditions, respectively. Greater response during auditory learning trials compared with cue and auditory–motor learning trials appeared in bilateral STG (Auditory > Cue) and bilateral IP (Auditory > AM). Greater response during motor learning trials compared with cue learning trials (Motor > Cue) appeared in left M1, bilateral S1, right S2, PMC, SMA, left SP, the cerebellum, as well as STG. Finally, compound contrasts between auditory and motor learning trials further demonstrated auditory and motor cortical response associated with auditory and motor learning trials, respectively. Right STG response was greater during auditory compared with motor learning trials ((Auditory > Cue) > (Motor > Cue) and (Auditory > AM) > (Motor > AM)). Response in a number of motor and premotor regions was greater during motor compared with auditory learning trials, including left M1, S1, and

Table 2. Pianists: BOLD Response Decrease during Learning and Recall Trials

Contrast	Region	Learning Trials		Recall Trials	
		Z	(x y z)	Z	(x y z)
<i>Decrease across All Blocks</i>					
Auditory > Cue	Right STG	3.12	(62 -14 14)		
Auditory > AM	Right STG	3.5	(48 -4 -10)		
	Right vPMC	3.35	(46 12 24)		
	Right IP	3.49	(66 -14 18)		
	Right S1			3.19	(34 -24 40)
	Left cerebellum			3.13	(-14 -52 -30)
	Left VC			3.01	(-14 -66 6)
<i>Decrease from Early to Late Blocks</i>					
Auditory > Cue	Right STG	3.60	(58 -12 8)		
	Right S2	3.63	(60 -12 22)		
Auditory > AM	Right STG	3.50	(52 -12 4)		
	Right S1	3.44	(62 -16 18)	2.86	(50 -12 32)
	Right vPMC	2.49	(58 8 14)		
	Left SP	3.48	(-12 -64 50)		
	Right SP	2.99	(18 -58 52)		
	Left cerebellum	3.49	(-26 -60 -26)	3.27	(-16 -46 -30)
	Right cerebellum	3.51	(20 -60 -26)		
	Right cerebellum	3.49	(30 -64 -54)		
	Right VC	4.00	(24 -78 22)		
	Left VC	3.52	(-28 -70 18)		
	Temporal-occipital	3.96	(-40 -46 -20)	2.82	(-56 -52 -12)

Coordinates for peak activations are in MNI-152 space, and peak z values are significant at $p < .05$, corrected. Cue = cue-only condition; AM = auditory-motor condition; M1 = primary motor cortex; S1 = primary somatosensory cortex; S2 = secondary somatosensory cortex; VC = visual cortex.

PMC ((Motor > Cue) > (Auditory > Cue)) and left M1 ((Motor > AM) > (Auditory > AM); Figure 5; Table 3).

Across all recall trials for nonmusicians, a similar frontal-temporal-parietal network responded in each condition compared with silence, including S1, SP, PMC, SMA, and STG. No response differences were detected between conditions during recall trials.

Nonmusicians: Decrease in BOLD Response over Learning and Recall Trials

Across all blocks and across early and late blocks of learning trials, BOLD response decreased in most conditions compared with silence. Across all blocks, response decreased during learning trials in right PMC during motor learning and in bilateral SP and right PMC during cue learning. Across early to late blocks, response decreased

mainly in ACC during auditory-motor, auditory, and motor learning trials, as well as in IP during cue learning trials. Contrasts between conditions mainly showed greater decreases during motor learning trials. Greater decrease in occipital cortex occurred during motor learning trials compared with cue and auditory-motor learning trials (Motor > Cue, Motor > AM).

BOLD response also decreased over recall trials, both across all blocks and from early to late blocks, in most conditions compared with silence. Across all blocks, response decreased in left IP and left middle temporal gyrus during motor recall and in left IP and occipital cortex during cue recall. Across early to late blocks, response decreased in SP, IP, and PMC during auditory-motor recall and in IP and temporal-occipital cortex during cue recall. No response differences were detected between conditions during recall trials.

Table 3. Nonmusicians: Mean BOLD response during Learning Trials

Contrast	Region	Learning Trials	
		Z	(x y z)
Auditory > Cue	Right STG	5.5	(48 -8 -2)
	Left STG	5.74	(-40 -26 6)
Motor > Cue	Left M1	4.76	(-36 -30 54)
	Left S1	5.56	(-44 -24 54)
	Left PMC	5.89	(-34 -22 66)
	Left SP	4.19	(-30 -48 66)
	Left STG	5.50	(-54 -20 10)
	SMA	4.64	(-2 -8 58)
	Right S1	6.15	(56 -16 42)
	Right PMC	5.05	(32 -10 70)
	Right S2	4.01	(58 -18 14)
Auditory > AM	Right cerebellum	5.17	(18 -64 -50)
	Left IP	2.52	(-32 -80 42)
	Right IP	4.62	(46 -80 26)
Auditory (>Cue) > Motor (>Cue)	Right STG	6.21	(54 -22 6)
Motor (>Cue) > Auditory (>Cue)	Left M1	5.6	(-32 -26 54)
	Left S1	5.75	(-38 -24 56)
	Left PMC	5.63	(-32 -22 66)
Auditory (>AM) > Motor (>AM)	Right STG	6.05	(68 -10 4)
Motor (>AM) > Auditory (>AM)	Left M1	5.95	(-34 -26 56)

Coordinates for peak activations are in MNI-152 space, and peak z values are significant at $p < .05$, corrected. Cue = cue-only condition; AM = auditory-motor condition; M1 = primary motor cortex; S1 = primary somatosensory cortex; S2 = secondary somatosensory cortex.

Nonmusicians: Increase in BOLD Response over Learning and Recall Trials

Across all blocks and across early to late blocks of learning trials, BOLD response also increased in each condition compared with silence. Across all blocks, response increased in parietal regions and in the cerebellum during motor learning trials. Across early to late blocks, response increased in occipital cortex during auditory-motor learning trials, occipital cortex and the cerebellum during cue learning trials and occipital, temporal, and PMC and the cerebellum during auditory learning trials. In addition, contrasts between conditions revealed greater response increases associated with both auditory and motor learning trials. Across all blocks, response increased more during motor learning trials compared with cue

learning trials in S1 and the cerebellum (Motor > Cue). Across early to late blocks, response increased more during auditory learning trials compared with cue and auditory-motor learning trials in left STG and the cerebellum (Auditory > Cue) and in left STG, S2, IP, and SMA (Auditory > AM) (Figure 7; Table 5).

BOLD response also increased over recall trials, across all blocks and across early to late blocks, in most conditions compared with silence. Across all blocks, response increased more during motor recall compared with cue recall in S1 and the cerebellum (Motor > Cue). Across early to late blocks, response increased more during auditory recall compared with cue recall in IP and occipital cortex (Auditory > Cue). Also across early to late blocks, response increased more during motor recall compared with cue recall in the cerebellum (Motor > Cue) (Table 4).

Nonmusicians: BOLD-Performance Correlations

To further examine the relationship between brain response and performance improvements during auditory learning, ROIs in IP and STG were selected for comparison with performance improvements. Similar to pianists, IP response was of interest as a multimodal region engaged more during listening compared with listening plus performance. The ROI in left IP was selected from peak activations from the contrast showing greater IP response during auditory compared with auditory-motor learning trials. Two additional ROIs in left STG were selected from peak activations from the contrast showing greater STG response during auditory compared with cue learning trials. Left IP response during learning and recall trials in the auditory condition (contrasted with the auditory-motor condition) was correlated with pitch and rhythm accuracy improvement (AUC) across recall trials for each nonmusician. Response in IP during auditory learning trials was negatively correlated with pitch improvement during recall trials ($r = -.59$, $p = .0153$), whereby nonmusicians who showed smaller average BOLD response in IP during auditory learning showed greater pitch improvement (Figure 6). Response in left STG during auditory learning trials (contrasted with the auditory-motor condition) was also negatively correlated with pitch improvement during recall trials, albeit more weakly ($r = -.46$, $p = .07$).

Nonmusicians: Functional Connectivity

To further probe how the auditory condition influenced auditory-parietal-motor connectivity, ROI regions in left IP and left STG were chosen as seed regions for functional connectivity analyses. Correlations with these regions in the auditory condition during learning and recall trials across the whole brain were examined, though none reached significance.

Table 4. Nonmusicians: BOLD Response Increase during Learning and Recall Trials

Contrast	Region	Learning Trials		Recall Trials	
		Z	(x y z)	Z	(x y z)
<i>Increase across Blocks</i>					
Motor > Cue	Left S1	3.30	(-52 -12 26)		
	Right S1			3.32	(50 -4 22)
	Left cerebellum	3.27	(-32 -64 -26)	3.28	(-10 -66 50)
<i>Increase from Early to Late Blocks</i>					
Auditory > Cue	Left STG	2.87	(-63 -18 6)		
	Right cerebellum	3.29	(24 -80 -46)		
	Left IP			3.10	(-56 -52 44)
	Left VC			2.50	(-50 -76 8)
Motor > Cue	Left cerebellum			2.53	(-32 -68 -46)
Auditory > AM	Left STG	2.96	(-44 -24 8)		
	Right S2	3.43	(52 -2 14)		
	Right IP	3.42	(52 -26 34)		
	SMA	3.4	(-12 -8 42)		

Coordinates for peak activations are in MNI-152 space, and peak z values are significant at $p < .05$, corrected. Cue = cue-only condition; AM = auditory-motor condition; M1 = primary motor cortex; S1 = primary somatosensory cortex; S2 = secondary somatosensory cortex; VC = visual cortex.

Conjunctions and Contrasts: Pianists and Nonmusicians

Table 5. Conjunctions: Pianists and Nonmusicians

Contrast	Region	Learning	
		Z	(x y z)
Auditory > Cue	Right STG	5.37	(56 -4 2)
	Left STG	4.56	(-62 -14 6)
Motor > Cue	Left M1	5.56	(-38 -22 58)
	Left S1	5.56	(-44 -24 54)
	Left PMC	5.51	(-32 -10 62)
	Left SP	3.64	(-22 -48 70)
	Left STG	5.50	(-54 -20 10)
	SMA	4.65	(-2 -8 58)
	Right S1	5.42	(50 -26 48)
	Right PMC	4.61	(34 -8 62)
	Right S2	4.58	(62 -22 20)
Auditory > AM	Left IP	3.76	(-34 -76 42)
	Right IP	3.85	(44 -74 38)

Coordinates for peak activations are in MNI-152 space, and peak z values are significant at $p < .05$, corrected. Cue = cue-only condition; AM = auditory-motor condition; M1 = primary motor cortex; S1 = primary somatosensory cortex; S2 = secondary somatosensory cortex.

To examine common activation patterns between pianists and nonmusicians, conjunctions were performed for the auditory and motor conditions, contrasted with the auditory-motor and cue conditions, for learning and recall trials separately. During learning trials, pianists and nonmusicians both engaged bilateral STG (Auditory > Cue), bilateral IP (Auditory > AM), and left M1, bilateral S1 and PMC, left SP, left STG, right S2, and SMA (Motor > Cue) (Table 5). A follow-up analysis contrasted response in these regions of common activation between pianists and nonmusicians. ROIs were defined based on peak voxels from clusters in left M1, left and right STG, and left and right IP. Average percent change in BOLD response across learning trials was extracted for each of these ROIs and compared between groups for the Auditory > Cue, Motor > Cue, and Auditory > AM contrasts. A 2 (group: pianists and nonmusicians) by 5 (ROI) by 3 (contrast) ANOVA revealed main effects of ROI and contrast and interactions between ROI and contrast but no main effects of group and no interactions between group, ROI, and contrast ($F_s < 1.9$, $p_s > .05$, $\eta^2_s < 0.021$).

DISCUSSION

This study compared the contribution of perceptual learning to that of motor learning in sensorimotor skill

acquisition for experts and novices. We tested whether pianists showed greater improvements and response in motor regions when learning to perform melodies by listening compared with performing, and we tested whether nonmusicians showed greater improvements and response in motor regions when performing compared with listening. As hypothesized, pianists demonstrated greater improvements in pitch and rhythm accuracy when learning by listening compared with performing. Contrary to our hypothesis, nonmusicians also showed greater improvements in rhythm accuracy, though not pitch accuracy, when learning by listening compared with performing. Overall, pianists and nonmusicians engaged a similar auditory–parietal–premotor network while learning by listening or performing and while performing from memory (recall). Pianists demonstrated learning-related decreases in premotor and parietal regions, specifically during auditory learning. In contrast, nonmusicians tended to show learning-related increases in premotor, parietal, auditory, and somatosensory regions during auditory learning, as well as increases in cerebellum and somatosensory response during motor learning. On average over learning trials, auditory regions (STG) were more responsive during auditory learning, whereas motor regions (M1/S1) were more responsive during motor learning. When pianists recalled melodies learned by listening, they showed greater activity in M1 and S1 across all recall trials compared with when they recalled melodies learned by performing. Activity in M1 correlated with activity in auditory and premotor regions during auditory learning, and M1 activity during auditory learning predicted pitch accuracy improvements at recall in the auditory learning condition. This pattern of results demonstrates an advantage for auditory over motor learning in skilled musicians that likely results from the capacity to transform auditory targets to motor commands via a feedforward control system (Tourville & Guenther, 2011; Guenther, 2006). Nonmusicians, like pianists, showed greater IP response during auditory compared with auditory–motor learning. Response in this region during auditory learning predicted improvements in pitch accuracy at recall during auditory learning. For nonmusicians, parietal cortex may have enabled the translation of pitch patterns to spatial coordinates for movement (Mattingley, Husain, Rorden, Kennard, & Driver, 1998; Gnadt & Andersen, 1988).

Overall, our results suggest both general and expertise-dependent advantages for auditory over motor learning. Skilled performers show an advantage for auditory learning and an enhanced motor cortical response during recall based on auditory learning. Furthermore, activity in M1 during auditory learning is related to improvements in performance across trials and is correlated with activity in posterior auditory regions. Based on these findings, we argue that this advantage stems from a trained auditory-to-motor network that facilitates learning through feedforward control. In contrast, a general advantage for

auditory learning was demonstrated by both pianists and nonmusicians. This may stem from a number of potential sources. First, it may arise from general-purpose auditory-to-motor transformation mechanisms that are not dependent on specific training. These mechanisms are likely instantiated in the auditory–parietal–motor network that underlies perception and production of both speech and music. In addition, auditory information may be easier to encode because it provides a target with relatively fewer degrees of freedom, that is, pitch height over time, in comparison with motor information, that is, joint angles over time. Finally, auditory learning was particularly advantageous for movement timing, likely due to an apparent advantage of sound for conveying temporal patterns (Repp & Penel, 2004).

Pianists demonstrated greater accuracy improvements and greater response in left M1/S1 during recall of melodies learned by listening compared with those learned by performing. This result suggests an advantage for auditory learning that depends on the capacity to transform auditory information into motor commands. It is hypothesized that, through training, musicians acquire a feedforward control system (Tourville & Guenther, 2011) linking auditory goals to the motor programs required to produce them. Evidence for auditory-to-motor transformation is well documented in musicians (Brown et al., 2013; Chen, Penhune, & Zatorre, 2008b; Baumann et al., 2007; Repp & Knoblich, 2007; Bangert et al., 2006; D’Ausilio et al., 2006; Finney & Palmer, 2003) and is thought to be underwritten by an auditory-to-parietal-to-premotor network (Frey, Campbell, Pike, & Petrides, 2008; Saur et al., 2008; Hickok & Poeppel, 2007; Zatorre et al., 2007). This transformation is thought to be enabled by a feedforward system, similar in principle to an inverse model, which can control action execution without feedback (Tourville & Guenther, 2011; Guenther, 2006). This system is thought to involve sensory–motor associations encoded by PMC and planned actions encoded by M1 (Tourville & Guenther, 2011). Under this model, greater M1 response in the context of auditory learning may arise because hearing the correct auditory outcomes translates to a more complete or accurate action plan than does practicing movements. In agreement with this, previous findings showed that M1 response increased with greater salience of a movement target (Seidler, Noll, & Thiers, 2004). In addition, M1 response increased over days of motor sequence training (Steele & Penhune, 2010), and M1 response patterns distinguished learned from novel motor sequences (Wiestler & Diedrichsen, 2013). These results further suggest that with training M1 helps to link movement goals to their correct execution. The correlation between greater pitch accuracy improvements and less M1 response in the auditory condition also corroborates this interpretation. Because M1 activity did not decrease across learning trials, this result points to a possible efficiency effect for experts (Lotze, Scheler, Tan, Braun, & Birbaumer, 2003). In summary, our results suggest

that pianists' ability to engage feedforward systems enables more efficient learning by listening compared with performing.

An advantage for auditory learning is further suggested by the changes in neural response over trials for both learning and recall. For pianists in all learning conditions, response in a bilateral premotor–parietal network decreased across learning trials. This network has been associated with sensorimotor mapping and memory retrieval (Kung, Chen, Zatorre, & Penhune, 2013; Chen et al., 2008b; Kostopoulos & Petrides, 2008; Zatorre et al., 2007). The auditory condition was associated with the most extensive decreases across learning trials, not only in auditory cortex but also in sensorimotor and frontal regions. These results suggest that skilled performers increase their auditory–motor mapping efficiency to a greater extent during learning by listening compared with performing with auditory feedback. In contrast to these parietal and premotor decreases over learning trials, the response in M1 associated with melody recall during auditory learning did not decrease over learning. Rather, M1 response remained consistent across learning. We posit that parietal–premotor decreases during auditory learning reflect an increase in sensorimotor mapping efficiency while yielding a consistent primary motor response over recall trials. Skilled performers may require consistent primary motor resources for execution but fewer sensorimotor integration resources as learning progresses.

When pianists were learning by listening, M1 response correlated with an auditory–motor network including STP, STS, and vPMC. Area STP has been proposed to be an auditory–motor interface (Hickok & Poeppel, 2007; Hickok, Buchsbaum, Humphries, & Muftuler, 2003) that has been found to respond equally robustly to perception and production of speech or music (Hickok et al., 2003), as well as to alterations in auditory feedback during speech or music (Pfordresher, Mantell, Brown, Zivadinov, & Cox, 2014; Tourville et al., 2008; Hashimoto & Sakai, 2003). STP may also serve as an “error map” in the feedback control of speech, comparing auditory output to the intended target (Tourville & Guenther, 2011). In the current study, pianists may have engaged this region while listening to acquire an auditory–motor mapping. Another region in this network, the STS, has been linked to higher-level auditory processing, such as voice recognition (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), speech perception (Möttönen et al., 2006), and the integration of sounds with other modalities (Vander Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009; Beauchamp, Yasar, Frye, & Ro, 2008; Hocking & Price, 2008). This region may enable the perception of meaningful units within the melodies, namely the pitch or interval identities, and it may work closely with area STP to map pitches or intervals to motor coordinates. Finally, vPMC has been linked to action planning, particularly in tasks where there is a one-to-one mapping between the sensory and the motor output (Chen et al.,

2012; Hoshi & Tanji, 2007). This region may have been involved in assembling the individual movements associated with each pitch into the complete action required to play the melodic sequence. This temporal–parietal–premotor network is analogous to networks proposed for speech production (Hickok & Poeppel, 2007) and motor control in speech learning (Tourville & Guenther, 2011). Pianists may utilize this network in concert with M1 while listening to detect, decode, and sequence the actions signaled by the auditory input, which in turn become the responsibility of M1 to execute when performing from memory.

Nonmusicians showed greater rhythm accuracy improvement when listening compared with performing. Listening may benefit rhythm learning for several reasons. First, people are generally better able to synchronize their movements with auditory stimuli than with visual stimuli (Hove, Fairhurst, Kotz, & Keller, 2013; Repp & Penel, 2004; Chen, Repp, & Patel, 2002), possibly due to the temporal precision offered by sound. Second, learners may have found it easier to reduce the degrees of freedom required to encode rhythms in the auditory modality compared with the motor modality (Nakahara et al., 2001). Finally, the human motor system is sensitive to temporal stimuli regardless of musical training. The current results showed that both pianists and nonmusicians showed greater rhythm accuracy improvements while listening compared with performing. Previous results have shown that nonmusicians engage motor regions of the brain while listening to rhythms or to sounds that follow a regular beat (Chen et al., 2008a, 2008b; Grahn & Brett, 2007; Chen, Zatorre, & Penhune, 2006). Nonmusicians may be capable of transforming auditory temporal information into movement via a feedforward motor control system that is better developed for the temporal domain than the pitch domain.

Nonmusicians showed greater IP response while learning by listening compared with performing with auditory feedback. Response in this region may reflect the transformation of sounds into spatial coordinates to guide motor production. The parietal cortex is thought to play a role in multimodal coordinate transformations (Foster & Zatorre, 2010; Zacks, 2008; Grefkes, Ritzl, Zilles, & Fink, 2004; Grefkes, Weiss, Zilles, & Fink, 2002) such as auditory–motor transformations in speech and music performance (Herholz, Coffey, Pantev, & Zatorre, 2016; Brown et al., 2013; Frey et al., 2008; Saur et al., 2008; Hickok & Poeppel, 2004, 2007; Zatorre et al., 2007). IP in particular has been linked to spatial perception (Andersen, Essick, & Siegel, 1985) and the ability to direct movements toward locations in space (Mattingley et al., 1998; Gnadt & Andersen, 1988). During auditory learning, parietal cortex may have encoded pitch patterns in terms of spatially directed movements. Nonmusicians who engaged this region less during auditory learning showed greater pitch accuracy improvements during recall. Because response in this region did not decrease

over trials, nonmusicians who engaged this region less may have improved more by encoding spatial patterns more quickly. A correlation between increased pitch accuracy and reduced parietal response was also previously shown by highly skilled pianists while playing melodies by ear (Brown et al., 2013).

Over the course of learning, nonmusicians tended to show increases in BOLD response as well as some decreases over learning. The most consistent increases occurred during motor learning in the cerebellum and somatosensory cortex. Response in the cerebellum also increased from early to late auditory learning (compared with cue-only learning) as did additional integration areas such as parietal and PMC (compared with auditory–motor learning). These increases could reflect the establishment of an auditory-to-motor mapping as learning progressed, in contrast to pianists for whom this was already in place. Increases in cerebellar response over motor skill learning may reflect movement optimization (Steele & Penhune, 2010). Increases in premotor response could reflect the establishment of auditory–motor associations, which is consistent with evidence for a premotor role in encoding pitch-movement associations during melody learning (Lega, Stephan, Zatorre, & Penhune, 2016). In addition, auditory learning produced greater increases than auditory–motor learning, perhaps due to the ability for nonmusicians to transform some features of sound into movement in a feedforward manner.

Both auditory and motor imagery likely played roles in melody learning for both pianists and nonmusicians. For both groups, auditory cortex responded during learning conditions without auditory stimulation, the motor and cue-only learning conditions. Auditory imagery has been reported for musicians during performance and for nonmusicians during musical memory tasks (Herholz, Halpern, & Zatorre, 2012; Halpern, Zatorre, Bouffard, & Johnson, 2004; Lotze et al., 2003). Nonmusicians also tend to engage motor-related regions such as the SMA when imagining melodies (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009), suggesting that auditory imagery may be related to active rehearsal, similar to the concept of the phonological loop (Baddeley & Andrade, 2000). Auditory imagery may also help pianists perform without auditory feedback (Brown & Palmer, 2013). Additionally, auditory and motor imagery may play roles in feedforward control of movement. Recent evidence shows enhanced corticospinal activity when nonmusicians hear the notes of a melody they have learned to play (Stephan, Lega, & Penhune, 2018), indicating that auditory information can predictively cue the associated motor response. In addition, visual information or imagery may have played a role. Visual cues may have helped support learning, because both auditory-only and cue-only learning involved the presentation of a complete and accurate target and because learning by ear can be difficult for nonmusicians (Engel et al., 2012). Because

listening to auditory targets facilitated learning relative to only viewing the visual targets, we posit that auditory learning facilitates auditory–motor skill acquisition relative to visual or motor learning.

It is important to note that the hypothesized advantage of perceptual compared with motor learning may only apply to early acquisition of a sensorimotor skill, as in the current study. Motor learning may be slower than auditory learning at first but may eventually catch up and surpass auditory learning by optimizing motor fluency and articulation. In the current study, auditory–motor learning was most difficult for both pianists and nonmusicians. However, auditory–motor learning could maximize learning in the long run via error correction. In accordance with this idea, professional musicians tend to learn new music by initially memorizing the overall structure followed by more intense practice of the motor execution (Chaffin, Lisboa, Logan, & Begosh, 2010; Chaffin & Logan, 2006). Visual–motor learning is also proposed to be dominated by perceptual learning at early stages and by motor learning at later stages (Nakahara et al., 2001). Thus, we interpret our results to show that a perceptual target facilitates early acquisition of a motor sequence, potentially through feedforward control. However, auditory–motor learning could be most favorable for optimization of performance (Chaffin et al., 2010) and even long-term memory for the music itself (e.g., auditory recognition; Brown & Palmer, 2012).

In summary, we demonstrate that learning via perception can be advantageous for both experts and nonexperts. The efficacy of perceptual learning at any level of expertise may depend on whether stimulus features can cue the motor system due to specific sensory–motor training, general experience, or intrinsic features of the sensory-to-motor processing stream. Sensorimotor learning at different levels of expertise could be conceptualized along a continuum of sensory-to-motor transformation capability or the degree to which an individual has developed feedforward control networks. To learn to perform a melody, beginners need to establish auditory–motor mappings, which may rely on parietal sound-to-space transformations. As learning proceeds, performance would be monitored and corrected by a feedback control system that adjusts production attempts and trains motor programs until they produce the correct target (Tourville & Guenther, 2011). Once auditory–motor mappings are established, the correct movements can be executed in a feedforward manner, which may be accomplished through an auditory–premotor–primary motor network (Tourville & Guenther, 2011). At this expert stage, further learning may be optimized by a feedforward system, in which auditory–premotor regions encode auditory targets and engage primary motor codes for execution. Both experts and nonexperts may also benefit from encoding low-dimensional auditory information due to greater temporal precision or an inherent link between auditory temporal information and movement.

APPENDIX

The image displays three sets of musical notation, labeled A, B, and C, each consisting of four staves.
Category A (Simple): Four staves of music in 4/4 time with a key signature of one sharp (F#). The melodies are simple, consisting of a few notes and rests.
Category B (Intermediate): Four staves of music in 4/4 time with a key signature of one flat (Bb). The melodies are more complex than category A, with more notes and some beamed eighth notes.
Category C (Complex): Four staves of music in 4/4 time with a key signature of two sharps (F# and C#). The melodies are the most complex, featuring many notes, including beamed eighth and sixteenth notes, and some triplets.

Figure A1. All melodies used in the experiment, notated in standard musical notation. (A) These melodies were categorized as “simple” and were learned by nonmusicians. (B) These melodies were categorized as “intermediate” and were learned by both nonmusicians and pianists. The notation depicted was presented to pianists during the experiment as the visual cue. (C) These melodies were categorized as “complex” and were learned by pianists. The notation depicted was presented to pianists during the experiment as the visual cue.

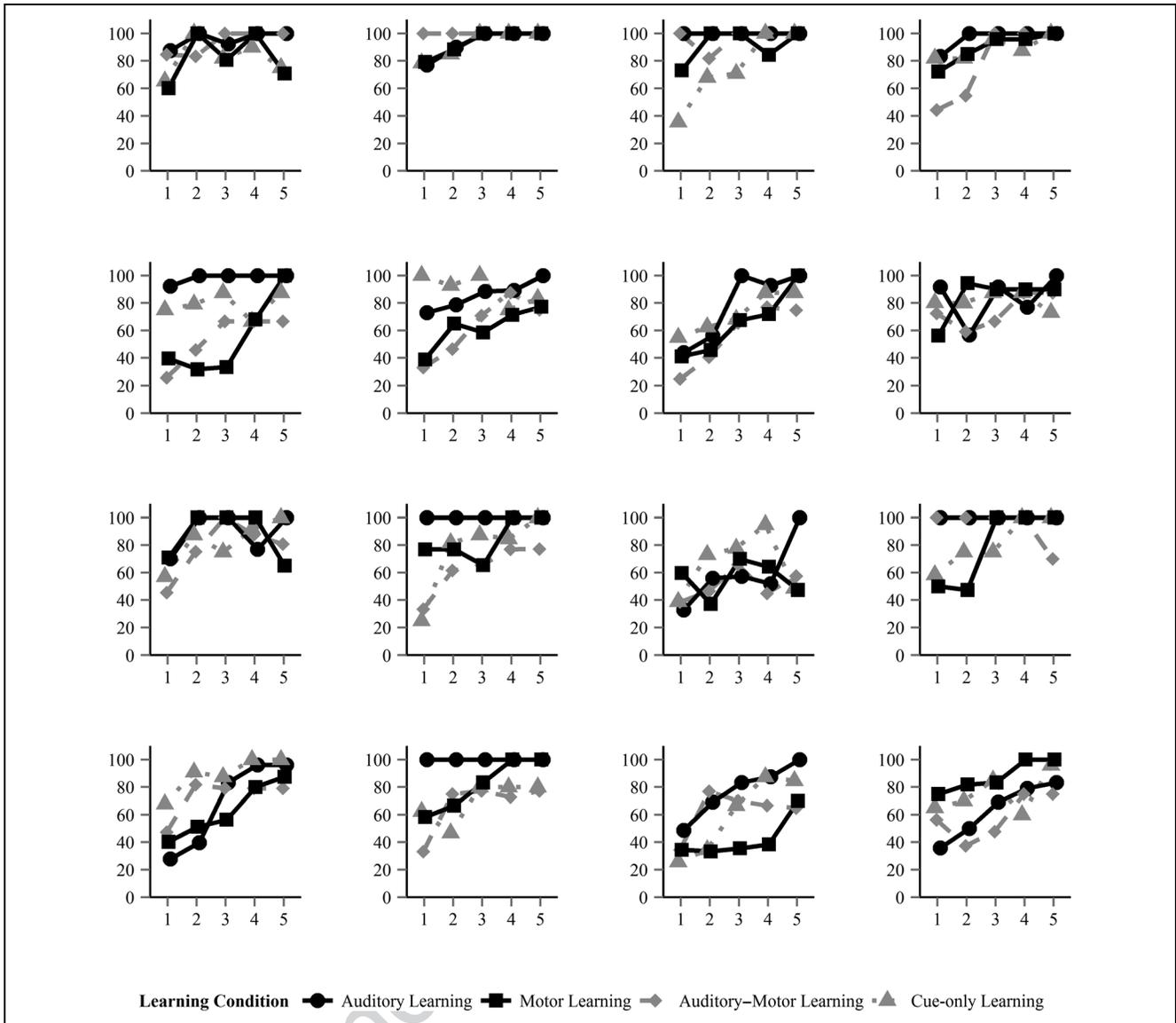


Figure A2. Pitch accuracy (all y-axes, percent correct pitches in the correct order) by recall trial (all x-axes) by learning condition for each individual in the pianist group.

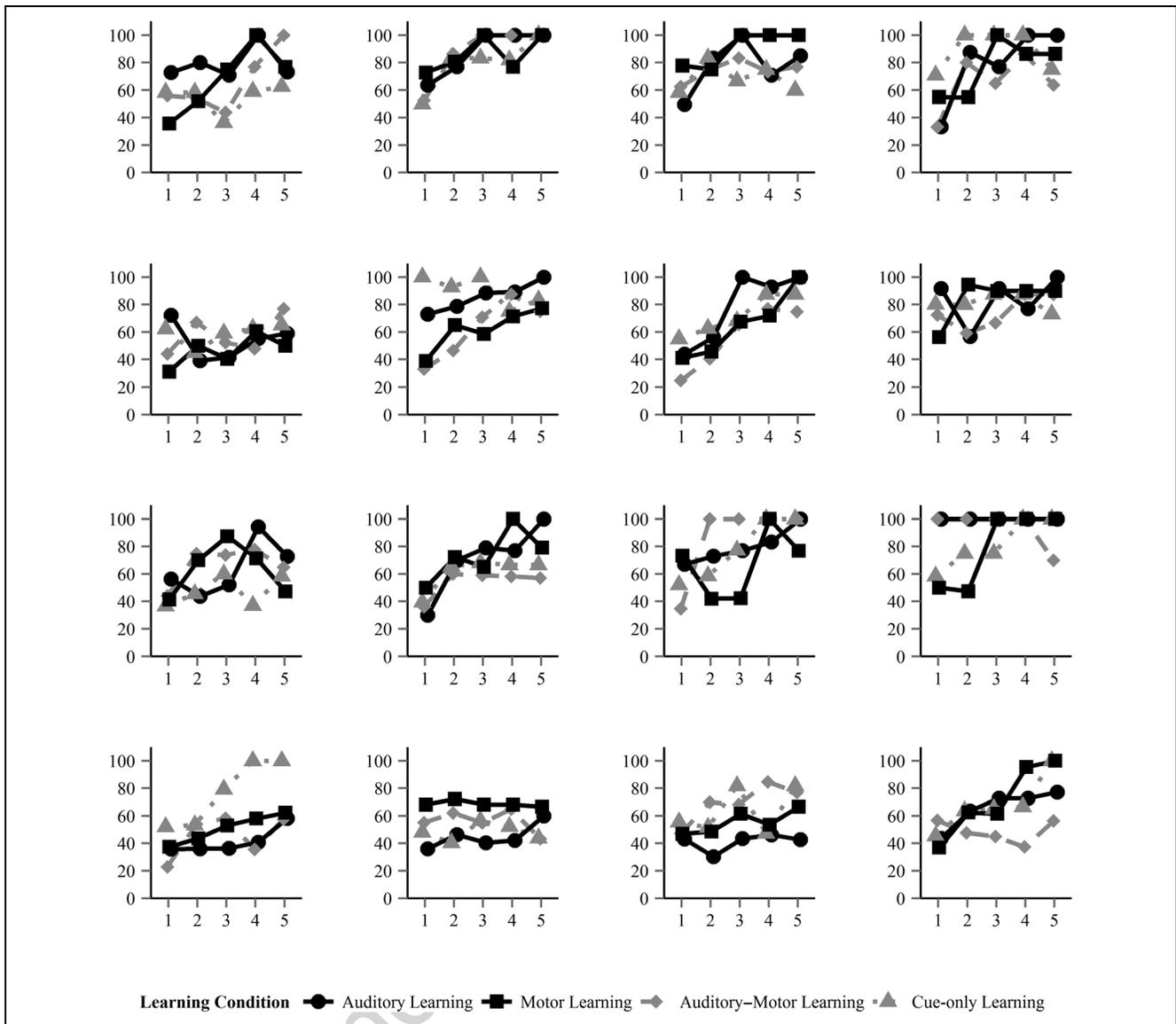


Figure A3. Pitch accuracy (all y-axes, percent correct pitches in the correct order) by recall trial (all x-axes) by learning condition for each individual in the nonmusician group.

Acknowledgments

This study was funded by the Canadian Institutes of Health Research (V. B. P.), the National Sciences and Engineering Research Council of Canada (NSERC-CREATE post-doctoral fellowship to R. M. B.), and the Quebec Bioimaging Network (scholarship to R. M. B.). We are extremely grateful to Joe Thibodeau for his programming expertise and template scripts, Avrum Hollinger for his wonderfully streamlined and expanded MR-compatible piano (Version 2) and for his help with the experimental setup, Spencer Rutherford for his invaluable assistance in conducting the experiments, and the Montreal Neurological Institute MR staff for running the MR sequences and coordinating the equipment setup so efficiently.

Reprint requests should be sent to Rachel M. Brown, Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Universiteitssingel 40, 6229 ER Maastricht, the Netherlands, or via e-mail: rachel.brown@maastrichtuniversity.nl.

REFERENCES

- Andersen, R., Essick, G., & Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, *230*, 456–458.
- Anguera, J. A., Russell, C. A., Noll, D. C., & Seidler, R. D. (2007). Neural correlates associated with intermanual transfer of sensorimotor adaptation. *Brain Research*, *1185*, 136–151.
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–145.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage*, *30*, 917–926.
- Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and nonmusicians. *Brain Research*, *1161*, 65–78.

- Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage*, *41*, 1011–1020.
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*, 417–429.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Brown, R. M., Chen, J. L., Hollinger, A., Penhune, V. B., Palmer, C., & Zatorre, R. J. (2013). Repetition suppression in auditory–motor regions to pitch and temporal structure in music. *Journal of Cognitive Neuroscience*, *25*, 313–328.
- Brown, R. M., & Palmer, C. (2012). Auditory–motor learning influences auditory memory for music. *Memory & Cognition*, *40*, 567–578.
- Brown, R. M., & Palmer, C. (2013). Auditory and motor imagery modulate learning in music performance. *Frontiers in Human Neuroscience*, *7*, 320.
- Brown, R. M., Zatorre, R. J., & Penhune, V. B. (2015). Expert music performance: Cognitive, neural, and developmental bases. In *Progress in brain research* (pp. 57–86). <http://doi.org/10.1016/bs.pbr.2014.11.021>.
- Chaffin, R., Lisboa, T., Logan, T., & Begosh, K. T. (2010). Preparing for memorized cello performance: The role of performance cues. *Psychology of Music*, *38*, 3–30.
- Chaffin, R., & Logan, T. (2006). Practicing perfection: How concert soloists prepare for performance. *Advances in Cognitive Psychology*, *2*, 113–130.
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*, 2844–2854.
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Moving on time: Brain network for auditory–motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, *20*, 226–239.
- Chen, J. L., Rae, C., & Watkins, K. E. (2012). Learning to play a melody: An fMRI study examining the formation of auditory–motor associations. *Neuroimage*, *59*, 1200–1208.
- Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, *32*, 1771–1781.
- Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, *21*, 515–532.
- D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *The European Journal of Neuroscience*, *24*, 955–958.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*, 524–531.
- Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*, 1–32.
- Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage*, *36*, 511–521.
- Engel, A., Bangert, M., Horbank, D., Hijmans, B. S., Wilkens, K., Keller, P. E., et al. (2012). Learning piano melodies in visuo-motor or audio-motor training conditions and the neural correlates of their cross-modal transfer. *Neuroimage*, *63*, 966–978.
- Fekedulegn, D. B., Andrew, M. E., Burchfiel, C. M., Violanti, J. M., Hartley, T. A., Charles, L. E., et al. (2007). Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosomatic Medicine*, *69*, 651–659.
- Finney, S., & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & Cognition*, *31*, 51–64.
- Foster, N. E. V., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*, 1350–1359.
- Frey, S., Campbell, J. S. W., Pike, G. B., & Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *Journal of Neuroscience*, *28*, 11435–11444.
- Gaab, N., Gabrieli, J. D. E., & Glover, G. H. (2007). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Human Brain Mapping*, *28*, 721–732.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, *9*, 416–429.
- Gnadt, J. W., & Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research*, *70*, 216–220.
- Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*, 893–906.
- Grefkes, C., Ritzl, A., Zilles, K., & Fink, G. R. (2004). Human medial intraparietal cortex subserves visuomotor coordinate transformation. *Neuroimage*, *23*, 1494–1506.
- Grefkes, C., Weiss, P. H., Zilles, K., & Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: An fMRI study implies equivalencies between humans and monkeys. *Neuron*, *35*, 173–184.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*, 350–365.
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281–1292.
- Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, *20*, 22–28.
- Herholz, S. C., Coffey, E. B. J., Pantev, C., & Zatorre, R. J. (2016). Dissociation of neural networks for predisposition and for training-related plasticity in auditory–motor learning. *Cerebral Cortex*, *26*, 3125–3134.
- Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*, 1382–1397.
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*, 486–502.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*, 673–682.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, *18*, 2439–2449.

- Hollinger, A. (2008). *Design of fMRI-compatible electronic musical interfaces (Unpublished masters thesis)*. McGill University, Montreal.
- Hollinger, A., Steele, C., Penhune, V., Zatorre, R., & Wanderley, M. (2007). fMRI-compatible electronic controllers. In *Proceedings of the 2007 International Conference on New Interfaces for Musical Expression (NIME07)* (pp. 246–249). New York.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *The Behavioral and Brain Sciences*, 24, 849–878; discussion 878–937.
- Hoshi, E., & Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: Anatomical connectivity and functional properties. *Current Opinion in Neurobiology*, 17, 234–242.
- Hove, M. J., Fairhurst, M. T., Kotz, S. A., & Keller, P. E. (2013). Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *Neuroimage*, 67, 313–321.
- Japikse, K. C., Negash, S., Howard, J. H., & Howard, D. V. (2003). Intermanual transfer of procedural learning after extended practice of probabilistic sequences. *Experimental Brain Research*, 148, 38–49.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17, 825–841.
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5, 143–156.
- Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, 70, 387–403.
- Kirsch, W., & Hoffmann, J. (2010). Asymmetrical intermanual transfer of learning in a sensorimotor task. *Experimental Brain Research*, 202, 927–934.
- Kostopoulos, P., & Petrides, M. (2008). Left mid-ventrolateral prefrontal cortex: Underlying principles of function. *The European Journal of Neuroscience*, 27, 1037–1049.
- Kung, S.-J., Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *Journal of Cognitive Neuroscience*, 25, 401–420.
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27, 308–314.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–146). New York: Wiley.
- Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, 29, 2477–2485.
- Lega, C., Stephan, M. A., Zatorre, R. J., & Penhune, V. (2016). Testing the role of dorsal premotor cortex in auditory–motor association learning using transcranial magnetic stimulation (TMS). *PLoS ONE*, 11, 1–16.
- Lotze, M., Scheler, G., Tan, H.-R., Braun, C., & Birbaumer, N. (2003). The musician's brain: Functional imaging of amateurs and professionals during performance and imagery. *Neuroimage*, 20, 1817–1829.
- Mattingley, J. B., Husain, M., Rorden, C., Kennard, C., & Driver, J. (1998). Motor role of human inferior parietal lobe revealed in unilateral neglect patients. *Nature*, 392, 179–182.
- Mitra, S., Amazeen, P. G., & Turvey, M. T. (1998). Intermediate motor learning as decreasing active (dynamical) degrees of freedom. *Human Movement Science*, 17, 17–65.
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage*, 30, 563–569.
- Nakahara, H., Doya, K., & Hikosaka, O. (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences—A computational approach. *Journal of Cognitive Neuroscience*, 13, 626–647.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, 6, 1–55.
- Newell, K. M., Broderick, M. P., Deutsch, K. M., & Slifkin, A. B. (2003). Task goals and change in dynamical degrees of freedom with motor learning. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 379–387.
- Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, 11, 63–68.
- Pfordresher, P. Q., Mantell, J. T., Brown, S., Zivadinov, R., & Cox, J. L. (2014). Brain responses to altered auditory feedback during musical keyboard production: An fMRI study. *Brain Research*, 1556, 28–37.
- Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, 28, 916–931.
- Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, 18, 6–7.
- Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, 68, 252–270.
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences, U.S.A.*, 105, 18035–18040.
- Savion-Lemieux, T., & Penhune, V. B. (2010). The effect of practice pattern on the acquisition, consolidation, and transfer of visual–motor sequences. *Experimental Brain Research*, 204, 271–281.
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82, 225–260.
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25, 336–354.
- Seidler, R. D., Noll, D., & Thiers, G. (2004). Feedforward and feedback processes in motor control. *Neuroimage*, 22, 1775–1783.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17, 143–155.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23(Suppl. 1), S208–S219.
- Steele, C. J., & Penhune, V. B. (2010). Specific increases within global decreases: A functional magnetic resonance imaging investigation of five days of motor sequence learning. *Journal of Neuroscience*, 30, 8332–8341.
- Stephan, M. A., Lega, C., & Penhune, V. B. (2018). Auditory prediction cues motor preparation in the absence of movements. *Neuroimage*, 174, 288–296.
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26, 952–981.

- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*, 1429–1443.
- Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science: A Journal of the American Psychological Society/APS*, *20*, 771–777.
- Verwey, W. B. (1999). Evidence for a multistage model of practice in a sequential movement task. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1693–1708.
- Wiestler, T., & Diedrichsen, J. (2013). Skill learning strengthens cortical representations of motor sequences. *eLife*, *2013*, 1–20.
- Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, *28*, 366–375.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*, 1317–1329.
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage*, *21*, 1732–1747.
- Worsley, K. J. (2001). *Functional MRI: An introduction to methods* (pp. 251–270). <http://doi.org/10.1093/acprof>.
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*, 1–19.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547–558.

Uncorrected Proof