Neural network retuning and neural predictors of learning success associated with cello training

Indiana Wollman[a,b,c,1], Virginia Penhune[b,c,d], Melanie Segado[a,b,c], Thibaut Carpentier[e], and Robert J. Zatorre[a,b,c]

[a]Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada; [b]Center for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montreal, QC H3A 1E3, Canada; [c]International Laboratory for Brain, Music and Sound Research, Montreal, QC H2V 2J2, Canada; [d]Laboratory for Motor Learning and Neural Plasticity, Department of Psychology, Concordia University, Montreal, QC H4B 1R6, Canada; and [e]CNRS, Institut de Recherche et Coordination Acoustique/Musique, Sorbonne Université, UMR 9912 Sciences et Technologies de la Musique et du Son, 75004 Paris, France

The auditory and motor neural systems are closely intertwined, enabling people to carry out tasks such as playing a musical instrument whose mapping between action and sound is extremely sophisticated. While the dorsal auditory stream has been shown to mediate these audio–motor transformations, little is known about how such mapping emerges with training. Here, we use longitudinal training on a cello as a model for brain plasticity during the acquisition of specific complex skills, including continuous and many-to-one audio–motor mapping, and we investigate individual differences in learning. We trained participants with no musical background to play on a specially designed MRI-compatible cello and scanned them before and after 1 and 4 wk of training. Activation of the auditory-to-motor dorsal cortical stream emerged rapidly during the training and was similarly activated during passive listening and cello performance of trained melodies. This network activation was independent of performance accuracy and therefore appears to be a prerequisite of music playing. In contrast, greater recruitment of regions involved in auditory encoding and motor control over the training was related to better musical proficiency. Additionally, pre-supplementary motor area activity and its connectivity with the auditory cortex during passive listening before training was predictive of final training success, revealing the integrative function of this network in auditory–motor information processing. Together, these results clarify the critical role of the dorsal stream and its interaction with auditory areas in complex audio–motor learning.

audio-motor integration | dorsal auditory-to-motor pathway | musical predisposition | fMRI | functional connectivity

Sensorimotor skills are essential in a great variety of human everyday life activities, such as speaking, cooking, and playing sports. These skills are acquired earlier or later and with more or less ease depending on the activity and the individual, but they are all acquired through repeated practice. However, such intensive practice is not without consequences on the brain. Neurophysiological evidence has accumulated over the past decades to demonstrate that the function and structure of the human brain can be modified by skill acquisition (1, 2). Such experience-dependent changes in brain circuits are commonly referred to as "brain plasticity." However, the underlying mechanisms of how such plasticity develops over time and how the neural substrates that happen to be important for the acquisition of a specific skill vary from one individual to another still remain to be fully understood.

As an example of a human sensorimotor activity, music performance provides an exceptional variety of naturally occurring complexities to study multisensory and motor processes at the neural level (for reviews, see refs. 3 and 4) and thus has emerged as a valuable model to study brain plasticity in all its complexity (for reviews, see refs. 5–7). Learning to play a musical instrument involves learning to integrate several sensory systems and the motor system to achieve remarkable feats of performance.

Several previous studies focusing on keyboard-based auditory–motor learning have examined neural changes during passive listening to a stimulus that the participant had been trained to play (8–12), while others examined neural responses during mute playing, i.e., a motion-only task (13, 14). These studies consistently showed learning-induced auditory–motor coactivation during both auditory-only and motor-only tasks, revealing the effects of cross-modal interactions between the auditory and motor systems. In particular, the dorsal auditory-to-motor cortical pathway, which has been conceptualized as playing a part in online auditory–motor transformation (15, 16), was shown to be especially recruited in abstract auditory–motor mapping tasks (17) and particularly in playing a musical instrument (4, 8).

Whereas prior studies have tested keyboard performance, string instruments present a different set of challenges. When playing a string instrument, one needs to develop asymmetric manual dexterity, since the left and right hands control rhythm and pitch with different, albeit synchronized, gestures. The right hand manipulates the bow, while the left hand controls the shifts in the position and pressure of the fingers on the strings. This complex motor coordination is accompanied by a significant somatosensory processing of afferent vibrotactile feedbacks of performance that have been observed in both behavioral (18, 19) and neurophysiological (20, 21) studies. Ultimately, a string player needs to learn an extremely sophisticated mapping between action and sound. Not only is there more than one action that can lead to a given pitch (depending on where the hand is placed along

Significance

In sophisticated auditory–motor learning such as musical instrument learning, little is understood about how brain plasticity develops over time and how the related individual variability is reflected in the neural architecture. In a longitudinal fMRI training study on cello learning, we reveal the integrative function of the dorsal cortical stream in auditory–motor information processing, which comes online quickly during learning. Additionally, our data show that better performers optimize the recruitment of regions involved in auditory encoding and motor control and reveal the critical role of the pre-supplementary motor area and its interaction with auditory areas as predictors of musical proficiency. The present study provides unprecedented understanding of the neural substrates of individual learning variability and therefore has implications for pedagogy and rehabilitation.

the fingerboard, different fingers on different strings can produce the same note), but also, unlike keyboards, which have a discrete one-to-one correspondence between motor action (key press) and pitch output, the auditory–motor mapping in a string instrument is continuous (analog). Therefore, a critical component of string-instrument learning is developing the ability to perform rapid online pitch corrections, i.e., instantaneous sensorimotor adjustments, to play in tune. This feature, which is also present in vocal pitch control (22–24), makes string instruments a valuable model for the study of auditory sensorimotor integration and motor control. Importantly, using an instrument that relies on fine, online adjustments allows behavioral assessment of the quality of the integration between auditory and sensorimotor systems shown in the correction of pitch accuracy during performance.

Finally, there is much evidence of behavioral interindividual differences in the learning of a new and complex task such as music performance, and neurophysiological evidence has begun to show that preexisting individual differences in anatomical and functional properties of the brain may affect learning rate or attainment (for a review, see ref. 25). The auditory cortex (AC) plays a prominent role in musical learning, since there are clear increases in the evoked response (26–28) and structural changes (26, 29, 30) in the AC of musically trained individuals, and since individual differences in the magnitude of task-evoked blood oxygenation level-dependent (BOLD) response in the AC predict the learning rate in various contexts such as pitch-discrimination tasks (31, 32) or piano playing (10). However, the importance of other regions within the audio–sensorimotor network still needs to be elucidated, together with the role of the AC within this network, in particular in learning tasks where auditory–motor integration is very prominent (such as string instruments). In particular, this could help better explain the heterogeneity in learning outcomes among individuals.

In the present longitudinal study we used cello training as a model for brain plasticity. We trained 13 participants with no musical background to play on a specially designed MRI-compatible cello (33) twice a week for 1 month. With the use of this unique device we were able to scan participants with fMRI while playing and listening to simple sequences on the cello before and after 1 wk and 4 wk of training (Fig. 1 *A* and *B* and *SI Appendix*, Fig. S1). We focused specifically on how the neural substrates of action–perception coupling associated with such learning change as a function of skill acquisition, how individual variability is reflected in the neural architecture over the course of learning, and whether there are neural predictors of such variability. Overall, we hypothesized that better players would exhibit stronger recruitment of the brain regions involved in the dorsal auditory-to-motor stream as well as stronger functional connectivity (FC) between these regions both before training and as training progresses. Indeed, we assume that a better ability to generate mental transformations of acoustic input into motor representations should result in more accurate anticipatory action control during performance and, in turn, in more accurate playing.

Results

Behavioral Training Data.
We developed two behavioral indices related to the playing task to capture pitch and timing accuracy; both measures show that subjects' performance improved significantly over the training sessions (T1–T7) (Fig. 1*C*). The average pitch errors (in cents, with 100 cents equaling 1 semitone) per training session decreased continuously from 59.3 ± 4.8 cents on T1 to 29.7 ± 3.6 cents on T7 [repeated-measures ANOVA, $F(1,72) = 6.20, P < 0.001$], which means that participants learned to play more in tune. Moreover, tempo errors (average timing deviation) (*Methods*) decreased continuously from 240.9 ± 36 ms on T1 to 91.3 ± 11.1 ms on T7 [repeated-measures ANOVA, $F(1,72) = 11.08, P < 0.001$], which means that subjects learned to

play more on the beat. Both behavioral patterns further suggest that there was a fast learning stage early in training, in which the improvement was relatively big, followed by a slower stage in which further gains were smaller from one session to another.

Behavioral Data in MRI Sessions.
For each subject, global performance scores per scan derived from pitch and tempo errors (*Methods*) were used to represent individual achievement. The higher the score, the better was the performance inside the scanner. Fig. 1*D* shows the evolution of pitch errors, tempo errors, and performance scores between scans at week 1 (SCAN wk1) and week 4 (SCAN wk4). Mean pitch and tempo errors decreased between the two scans, but the decrease was significant only for pitch errors [repeated-measures ANOVA, $F(1,12) = 9.35, P = 0.009$ for pitch error; $P = 0.15$ for tempo errors]. Overall, subjects' performance scores improved between scans, with a mean 28.8% increase [repeated-measures ANOVA, $F(1, 12) = 4.1, P < 0.01$].

The relationships between pitch and timing errors were examined using Pearson correlation coefficients computed across participants. No correlations were found at $P < 0.05$ between the two indices in any training session or in SCAN wk1. In SCAN wk4, smaller pitch errors were associated with smaller timing errors ($r = 0.62, P < 0.05$), indicating that the two measures are rather independent until sufficient expertise has been acquired.

Task-Related Networks.
Contrast images for task vs. rest were computed to assess basic task-related activity at SCAN wk4 for the set of musical sequences that were trained. As a control, in the listen-only condition we used a sequence that was never trained but that had the same tones as the trained sequences but presented in a different order. The network associated with passive listening of the learned sequences after training engages bilateral primary and secondary auditory cortices (AC), parts of the motor network [the supplementary motor area (SMA) and dorsal premotor cortex (PMC)], and posterior parietal regions [the superior parietal lobule (SPL)] (*SI Appendix*, Fig. S2). The cello-playing network encompasses the auditory regions, the somatosensory regions, and the motor production network, i.e., cortical and subcortical motor regions, including primary motor (M1), premotor and supplementary motor areas (the PMC, SMA, and preSMA), the basal ganglia, and the cerebellum.

Training-Related Effects.
Because training sessions started only after the initial scan (preSCAN), changes in brain activation across short-term learning, i.e., after 1 wk of training (SCAN wk1 vs. preSCAN), could be investigated only in the passive-listening task (i.e., playing was not possible at preSCAN, since participants had not yet received any training). Changes in activation across the later learning phases (SCAN wk4 vs. SCAN wk1) were examined in the listening and playing tasks.

Already after 1 wk of training, and persisting after 4 wk, we found that the supplementary motor areas (preSMA and SMA), the right dorsal premotor cortex (dPMC), and the left posterior parietal cortex (SPL) responded more to the passive listening of the learned sequences than seen in the preSCAN (Fig. 2*A* and *SI Appendix*, Table S1*A*). In each of the three regions, significant differences in parameter estimate were also observed between preSCAN and SCAN wk1 [SMA: $t(12) = 3.51, P = 0.004$; dPMC: $t(12) = 3.88, P = 0.002$; SPL: $t(12) = 7.37, P < 0.001$, after Bonferroni correction] but not between SCAN wk1 and SCAN wk4 (all Ps > 0.4) (Fig. 2*A*, *Insets*). However, no significant differences were found within the bilateral AC across the three scan sessions (all Ps > 0.17). For the playing task, no training-related changes were observed between SCAN wk1 and SCAN wk4 at the statistical threshold, presumably because most of the changes had already occurred after 1 wk of training. A conjunction analysis
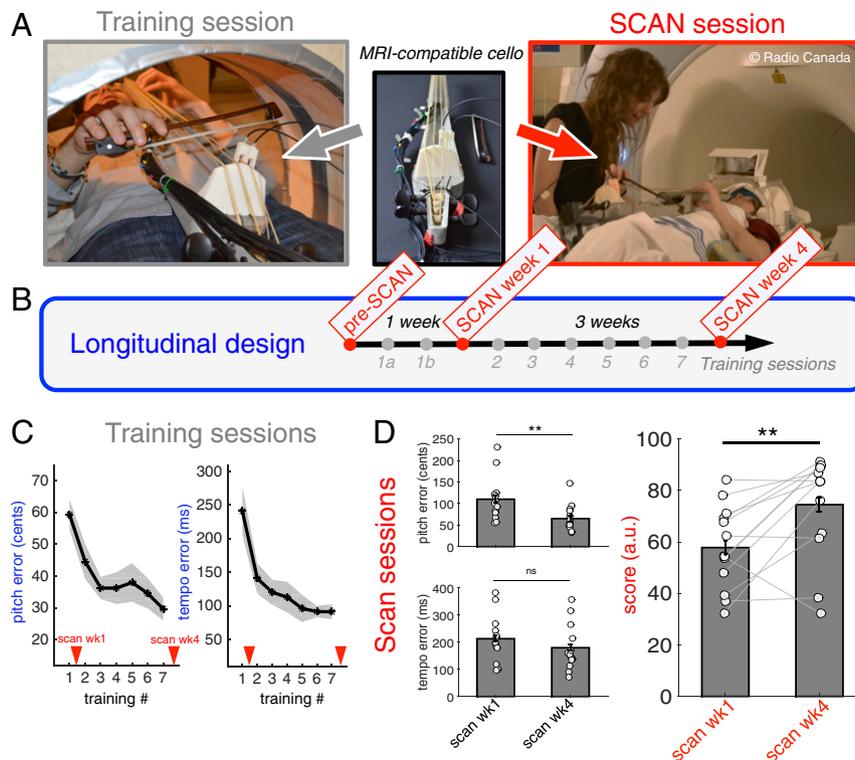
Fig. 1. Design and behavioral results. (*A*) An MRI-compatible cello was used for training and scan sessions. (*B*) The longitudinal repeated-measure protocol with three scans and eight training sessions. (*C*) Group average pitch errors and tempo errors across training sessions. (*D*) Group average pitch errors, tempo errors, and performance scores across scanning sessions. Here, the performance scores were linearly rescaled taking into account the pitch and tempo datasets of both scans (*Methods*). Error bars and shaded areas show SEM. **$P < 0.01$, paired *t* tests.

shows that the three regions that were more engaged during the listening task in SCAN wk4 relative to preSCAN (SMA, dPMC, and SPL) were also commonly active during the playing task at the last scan (SCAN wk 4) (Fig. 2*B*). Moreover, it is of note that there is a complete overlap in the SMA and dPMC and, to a lesser extent, in the SPL.

Based on these results, we assessed patterns of FC (*Methods*) between these three regions, and we also included the bilateral AC. We selected the left SPL as a seed because it appeared to have a key role in the listening task across learning (the overlap between the listening and playing tasks was only partial). FC analyses were conducted between the seed and a priori-defined regions of interest (ROIs) during the listening task in each scan. In the preSCAN, there was a positive correlation between the left SPL and bilateral AC [specifically, the planum temporale (PT)] and the SMA (*SI Appendix*, Fig. S3). After 4 wk of training, higher connectivity was found between the SPL and bilateral AC and the SMA (SCAN wk4 > preSCAN) (Fig. 2*C* and *SI Appendix*, Table S1*B*). Enhanced connectivity with the right dPMC reached a marginal level of significance ($P = 0.06$).

The same analyses were performed with the listening trials of the untrained sequence that do not present full motor significance, since the untrained sequence is composed of the same notes but in a different order and was never practiced. Changes in the fronto-parietal network were also observed across the training period for these trials, but the same FC analyses did not show the plastic changes detected for the trained sequence, indicating a degree of specificity to the reorganization (*SI Appendix*, Fig. S4).

Training-Related Effects and Individual Performance. To further test whether brain changes were modulated by the individual behavioral achievements, we performed regression analyses between

BOLD signal variation and the global performance score. We used the score at SCAN wk4 as a regressor. Results are presented in Fig. 3 and *SI Appendix*, Table S2. During passive listening to the trained sequences in SCAN wk4, participants with highest performance scores showed increased activity (relative to preSCAN) in the right superior temporal gyrus (STG), right putamen, and left middle frontal gyrus (extending into the left secondary somatosensory cortex). In the playing task, better performance at week 4 relative to week 1 was also positively correlated with signal changes in the right STG and additionally in the right hippocampus. In the playing task, the same analysis was conducted using only the pitch score or only the timing score as a regressor. Results show that increased activity in hippocampus correlated with both pitch and timing scores, whereas increased activity in STG correlated only with the pitch score (*SI Appendix*, Fig. S5). No significant cluster was observed for negative correlations.

Predispositions.

Pretraining listening task. Regression analyses were performed to test for correlations between the pretraining BOLD response during listening and training success after 1 month as defined by the performance score at SCAN wk4. Results in Fig. 4*A* show that better performance at the end of the training month was predicted by stronger activation of the preSMA during listening at preSCAN. We performed a robust correlation analysis (34) and detected no outlier (Fig. 4*B*). To further ensure the robustness of this finding, we employed a leave-one-participant-out cross-validation procedure (35) that allows defining ROIs with an independent dataset (*Methods*). Results are presented in *SI Appendix*, Fig. S6. This procedure allows a more unbiased estimate of the brain activity, to avoid circularity and give a better estimate of out-of-sample prediction. The results found when
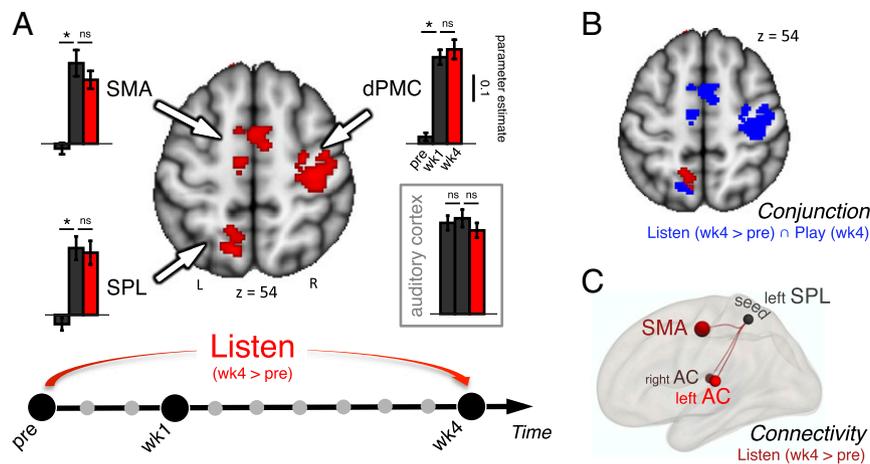
Fig. 2. Training-related changes. (*A*) Functional activation changes during passive listening to the learned sequences across the 4 wk of training (SCAN wk4 > preSCNA). Contrast images are displayed with a cluster-based thresholding (z >2.3) corrected for multiple comparisons (*P* < 0.05). Enhanced activations are observed in regions of the dorsal auditory-to-motor pathway including a left parietal region (SPL), SMA, and the dPMC. *Insets* show bar plots of BOLD signal within the three regions of the dorsal pathway (indicated by arrows) across the preSCAN, SCAN wk1, and SCAN wk4 scans (*P < 0.05. paired-sample *t* tests after Bonferroni correction). No significant differences (ns) in parameter estimates are observed in regions of the dorsal pathway between week 1 and week 4 suggesting that the sensorimotor integration loop is already efficient after 1 wk of training. As a reference, no significant changes are observed within bilateral AC across scans. (*B*) Conjunction Play week 4/Listen week 4 >pre (in blue) superimposed with functional activation changes (week 4 > pre) during listening (in red; z >2.3). (*C*) Between-scans contrast of FC (week 4 > pre) shows increased connectivity between the SPL and the a priori-defined ROIs: the bilateral AC (specifically, the PT) and SMA (*Methods*) (*P* < 0.05 FDR-corrected).

using this latter procedure were very similar to the original regression analysis, indicating that the finding is stable when tested with a more stringent cross-validation method. Also, the result was similar when either only the pitch score or only the timing score was used as a regressor (*SI Appendix*, Fig. S6*C*).

Based on this consistent regression result in the preSMA, we performed FC analyses using this functionally defined preSMA region as the seed region and the performance score at SCAN wk4 as a covariate to examine whether individual differences in FC involving the preSMA also correlated with performance achievements. Results showed that better performance at SCAN wk4 was associated with stronger connectivity between the preSMA and bilateral AC (specifically the STG) during listening at preSCAN (Fig. 5*A* and *SI Appendix*, Fig. S7*A*).

Pretraining resting state. Finally, we investigated whether any functional markers of a predisposition for cello learning could be detected during a resting-state scan obtained before training. FC analyses of pretraining resting-state networks were performed using the functionally defined preSMA region (see above) as the seed region and the score at SCAN wk4 as a covariate to test for correlations between the intrinsic resting-state FC of brain networks and training success after 1 month. Results show that better performance at SCAN wk4 was associated with stronger resting-state FC between preSMA and AC (specifically the PT) at preSCAN (Fig. 5*B* and *SI Appendix*, Fig. S7*B* and Table S3). This pattern of resting-state FC partially matches the one observed in the pretraining listening condition (Fig. 5*A* and *SI Appendix*, Fig. S7*A* and Table S3).

We also performed robust nonparametric correlation analyses (34) on these data. In the listening task, we detected no outlier, and the Spearman correlation between connectivity in AC/preSMA and score at SCAN wk4 is significant (r = 0.65, CI 0.17, 0.91). We also detected no outlier for the resting-state data, and the Spearman correlation between connectivity in the right AC/preSMA and the score at SCAN wk4 is significant (r = 0.57, CI 0.03, 0.87).

Discussion

The present study was designed to identify the changes in brain activity patterns and their relation to behavioral changes associated

with cello learning, i.e., changes with repeated practice of complex auditory–motor sequences. The main findings of this study are (*i*) the evidence of functional reorganization in the auditory–motor network, as indexed by the recruitment of dorsal–cortical stream regions after training and by changes in connectivity between the auditory and motor systems (Fig. 2); (*ii*) interindividual differences in neural plasticity associated with this learning, as shown by stronger recruitment of regions related to stimulus encoding (AC and hippocampus) in the better players (Fig. 3); and (*iii*) evidence of a functional predisposition for cello training success as shown by stronger recruitment of the preSMA (Fig. 4) and its connectivity to auditory regions (Fig. 5) before training in the better learners. These findings are discussed in the next paragraphs.

Auditory-to-Motor Dorsal Stream as a Canonical Neural Substrate Underlying Musical Training. From SCAN wk1 we identified an extended network related to the playing task that encompasses the auditory regions, the motor regions, the dorsal pathway, and the cerebellum, which is consistent with other studies of performance (36, 37). Our study further reveals that subcomponents of that network are jointly recruited in listening as a consequence of learning (Fig. 2*B*). Indeed, in line with previous research (9–11), we show that cello training induced increased activity in regions of the auditory-to-motor dorsal cortical pathway, including the preSMA and SMA, the right dorsal PMC, and parietal regions, during passive listening to the learned sequences (Fig. 2). Prior studies, however, have not compared listening conditions with playing in the scanner, to allow a direct comparison of the two, as shown in the conjunction analysis (Fig. 2*B*). We thus are able to show directly that many of the same structures within the dorsal stream are similarly activated during passive listening and music performance. Our study also used a string instrument, which requires action–sound mapping different from that with keyboard instruments, which have been the focus of previous research. Since similar networks appear to be active across studies, we conclude that the dorsal stream has a general capacity for action–sound mapping. This conclusion is also consistent with recent data showing that cello playing and singing engage similar brain networks (38). These findings
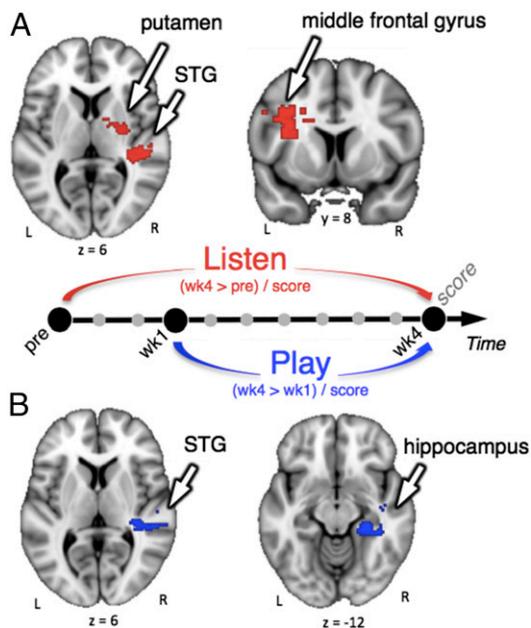
Fig. 3. Interindividual differences in the neural substrates of brain plasticity. Regression analyses with performance score at SCAN wk4 (with pitch and tempo errors rescaled taking into account their respective datasets in SCAN wk4) of BOLD signal variation in the listening task (week 4 > pre) (*A*) or in the playing task (week 4 > week1) (*B*). All images are cluster-corrected ($z > 2.3$, $P < 0.05$). During the listening task, significant effects were observed in left middle frontal gyrus (extending into the left secondary somatosensory cortex), right putamen, and right STG. During the playing task, significant effects were observed in right hippocampus and right STG.

therefore highlight the auditory–motor coupling (4, 16, 39, 40) that develops during music learning.

That no statistically significant training-related changes in brain activity were observed between SCAN wk1 and SCAN wk4 for the playing task, even though continued learning was observed in the behavioral measures (Fig. 1 *C* and *D*), could be due to the use of only four simple melodies of five notes each (our stimuli represent a compromise between complexity and learnability that allowed us to study longitudinal training effects based on performance levels). Had we used more difficult material it might have led to more range to pick up longer-term learning effects, although then we might not have been able to measure any reasonable performance at week 1.

On a more conceptual level, the dorsal stream can be seen as mediating complex internal models of cello performance, i.e., the brain's capacity to automatically trigger sensorimotor processes that resemble those associated with actual playing just by listening to the trained sounds, as proposed by Keller in the context of internal models (41). These mechanisms may support the generation of anticipatory images that may facilitate action planning (42). Our study reveals that these models are formed even at early stages of learning, since increased activity in the dorsal stream was already observed after only 1 wk of training, when performance is not yet highly practiced or accurate, and this activity did not evolve afterward (i.e., between week 1 and week 4), whereas behavioral performance continued to improve (Fig. 2*A*).

Among the regions of the dorsal pathway in which changes are observed during passive listening, the posterior parietal region was found to act as a key node in this neural retuning. Indeed, functional plasticity during listening is characterized by an increase in the FC between the posterior parietal cortex and both the auditory and premotor systems over the training month (Fig. 2*C*). Therefore, our data provide further evidence that the

parietal region is a key sensory–motor interface (39, 43) when perception is linked to learned action, a concept which is also well aligned with models of visuomotor function (44).

Interestingly, changes in the fronto-parietal network were also observed for the untrained sequence that was passively experienced and for which there was no direct motor representation (*SI Appendix*, Fig. S4). However, there were no changes in the connectivity pattern as seen for the learned sequences. These observations mark a difference in relation to prior studies using a keyboard where new melodies did not elicit any sensory–motor coactivation (10, 11). This difference could be attributed to the emphasis in the present study being placed on the flexible many-to-one mapping enabled by the use of a cello (we used sequences composed of identical pitches but differing in their fingering pattern). It has been suggested that a larger redundancy space, i.e., motor variability in exploring a new movement, determines faster motor learning (45). Hence, these changes could reflect the partial learning of a more general sensory–motor association; such associations have been shown to form even in the absence of training under some circumstances (46) and to involve the dPMC (47). It is also possible that the training resulted in a sort of generalized enhancement of attention to cello stimuli that are composed of the same notes but perhaps not to the extent of being able to play a new sequence.

Better Performers Optimize the Recruitment of Regions Involved in Auditory Encoding and Motor Control. Previous studies identified pretraining activity in the AC as a marker of predisposition for success in fine-grained pitch-pattern learning (32), linguistic pitch-contour learning (48), and, together with the hippocampus, in complex piano learning (10). Here we identified several cortical and subcortical regions, including right AC and hippocampus, whose increased activity induced by training correlated positively with performance in both listening and playing tasks (Fig. 3).

Our data thus provide further evidence that the AC, particularly in the right hemisphere, plays a major role in auditory–motor learning, and they confirm the enhanced role of the hippocampus in successful melodic memory retrieval across learning (10, 49). Here, instead of being predictors of performance achievement, these two regions reflected interindividual differences in neural plasticity and how fluctuations in neural activity across training correlate with performance. However, the nature of the task and the metrics taken as performance indicators are not quite comparable to those in previous studies, because our measures were direct indices of accuracy (that are sensitive to sensorimotor integration) and do not reflect the learning rate, as was the case in previous studies (10). Our data further demonstrate that more accurate performance relies on a finer-grained encoding of pitch information (vs. tempo information) in the right AC across learning, which is consistent with the hypothesis that the right AC is related to better spectral resolution (50).

Last, the greater recruitment of the putamen during listening in those subjects who played better is consistent with its role in beat and rhythm processing in perceptual tasks (51, 52), and increased activity in the middle frontal gyrus suggests an enhanced auditory working memory (53, 54). Overall this may reflect an increased sequence-specific priming of action representations in better players who put themselves more in an active (vs. passive) listening context across scanning sessions. Although our participants were instructed to listen passively to the sequences in the listening condition, our design is such that listening to a sequence always occurred before a playing trial. Thus, enhanced processing of sequences in the basal ganglia and frontal system during listening could lead to more accurate performance. It might also be that the development of optimal patterns of FC within these systems during training benefits both perception and action.
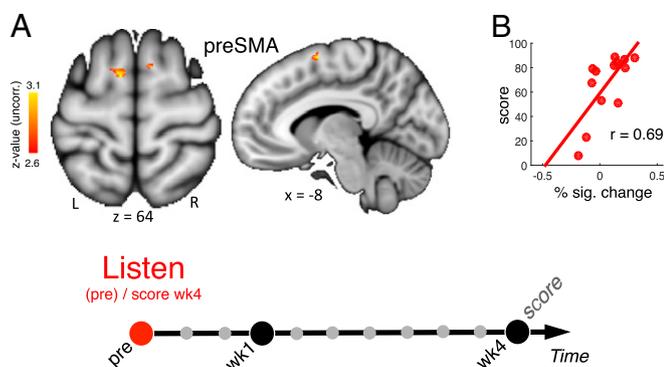
Fig. 4. Predispositions. (*A*) Regression analyses with behavioral variables measured at the end of the training (week 4) testing for correlations between pretraining activity during the listening task and training achievements (z >2.6, i.e., $P < 0.01$ uncorrected). (*B*) Robust Spearman correlation between percentage of BOLD signal change in preSMA at preSCAN during listening (*x* axis) and the performance score at SCAN wk4 (*y* axis) (CI 0.16, 0.94). Individual differences in pretraining preSMA activity during the listening task were predictive of training success.

Overall, our data provide further evidence that greater activity in areas of the cortico-basal ganglia circuit contributing to stimulus encoding and storage and hippocampal structures are essential for building the right memory trace necessary for motor performance accuracy (55).

PreSMA and Its Interaction with the Auditory System as Predictors of Musical Proficiency. In our study, individual differences in performance were not correlated with changes in neural activity within the dorsal stream over the training period. Rather, the increased engagement of the dorsal system with training appears to be common feature of cello learning and therefore could be viewed as a general prerequisite for musical instrument performance. However, our study reveals that the fluctuations in the initial state of the dorsal stream before training provide interesting information concerning the neural substrates underlying individual differences in learning. Indeed, the individuals who were better able to play the cello at the end of the training showed higher levels of activity in the region anterior to the supplementary motor area (the preSMA) during listening pretraining as well as higher levels of FC between the preSMA and bilateral AC during that same task and even between the preSMA and right AC during a resting-state scan (an independent dataset acquired before training) (Figs. 4 and 5), albeit in a somewhat more dorsal preSMA location than the preSMA/SMA area that showed training-related changes.

The preSMA region is considered part of the dorsal auditory stream and, along with the dPMC, is thought to play a role in linking sound and action (4, 16, 56). Previous studies showed that this region was involved in the acquisition and storage of new motor skills (57, 58) and in action execution (59, 60). In addition, the preSMA has been conceptualized as an audio–sensorimotor node, shaping the auditory perception of learned actions (for a review see ref. 56). A few other studies on sensorimotor expertise revealed that, compared with nonexperts, athletes recruit the preSMA more strongly when listening to their sport's sounds (61), and musicians show stronger activity in the preSMA when listening to music played on their instrument of expertise (62). Finally, stronger engagement of the preSMA is related to greater sensitivity in beat perception (63). The present findings extend these views, even within a group of nonexperts. Here, the predictive role of the preSMA and its interaction with the AC most likely reflect nonexperts' basic abilities to map sounds to action so that the subsequent performance of complex audio sensorimotor

sequences is facilitated. Moreover, that the preSMA predicted both pitch and temporal accuracy in performance (*SI Appendix,* Fig. S6) suggests that it relates to higher sensitivity to music features in the broad sense. This study opens the way to address new issues, such as deciphering the respective contribution of individuals' genetic background versus their previous experience, at the basis of this functional predictor. Overall, that the FC between motor and auditory systems before the training can be a determinant in learning potential supports the view that cello learning, first and foremost, requires tight auditory–motor coupling.

Conclusion
Our longitudinal study of the acquisition of a complex skill, cello playing, reveals dissociations between brain networks that undergo plastic modifications as a function of training, versus those that are most related to performance, versus those that serve as predictors of learning outcome. The circuits involved demonstrate different contributions to perceptual, motor, and cognitive functions depending on the aspect of learning probed. These findings would have been difficult to observe without the use of complex, realistic tasks coupled with longitudinal brain imaging and behavioral outcome measures; such approaches therefore hold promise for future research on ecologically valid sensory–motor learning, with potential eventual applications in clinical rehabilitation and pedagogy.

Methods
Participants. Thirteen right-handed healthy volunteers participated in the present study (six men; mean age, 26 ± 4 y; age range, 20–31 y). Seven participants were native French speakers; the other six were native English speakers or used English as their working language. Participants were selected for their lack of musical background. None had previously received any formal musical training (other than at primary school), none had experience with a string instrument, and none was currently engaged in active music making. To ensure they did not have musical deficit before the experiment, participants were screened on their global performance on the Montreal Battery of Evaluation of Amusia (64). The Local Ethics Committee from the Montreal Neurological Institute, McGill University approved the protocol, and the subjects gave their informed consent.

Longitudinal Design. Participants were trained to play the cello over the course of 1 mo and were scanned at three time points (Fig. 1*B*). Immediately after the first MRI session (preSCAN), participants began the 4-wk cello training period. The second scan (SCAN wk1) took place after 1 wk of
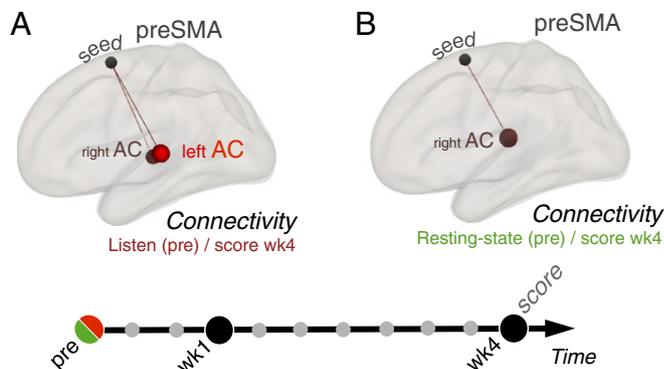
Fig. 5. Predisposition in FC analyses. (*A*) Listening task: regression analyses of pretraining FC (pre) with the performance score at last scan (wk4) shows a positive correlation between connectivity between the preSMA and bilateral AC (specifically, posterior STG; a priori-defined ROIs) (*Methods*) and training achievements ($P < 0.05$ uncorrected). (*B*) Resting state: regression analyses of pretraining FC with the score at the last scan shows a positive correlation between connectivity between the preSMA and right AC (specifically, the right PT; a priori-defined ROIs) (*Methods*) and training achievements ($P < 0.05$ uncorrected).

training, during which participants learned the basis of holding and playing the cello and were introduced to and began to practice the four musical sequences included in this training study. The last scan (SCAN wk4) took place at the end of the fourth week of training. The preSCAN and SCAN wk4 were scheduled exactly 4 wk apart. The training protocol was created for the purpose of this study and was piloted on three additional nonmusicians who were not included in the study.

MRI-Compatible Cello. Participants performed inside and outside the scanner on a specially designed MRI-compatible cello (33) (Fig. 1*A*). This instrument was made from nonconductive and nonferrous materials (fiberglass, epoxy resin, plastic, wood, and, for historical correctness, gut strings). The instrument and bow were miniaturized to enter the scanner. Two specialized optical sensors mounted perpendicular to the D and A strings, respectively, captured the acoustic performance. The resulting sound was transmitted to the player via headphones in real time and was recorded for sound performance analyses.

Stimuli. Participants learned to play four musical sequences spanning two cello strings, A and D, at 50 bpm (*SI Appendix*, Fig. S1*B*). All sequences are composed of five notes of a D major scale beginning on D3. Two types of tone sequences were included: a scale (consecutive pitches) and a melody (the same pitches but in a different order). In each type, two different fingering patterns were included: simple (no motor shift) or complex (a motor shift of the arm to accompany finger placement). One additional melody composed of the same five pitches but in a completely different order was presented during the scan sessions only to control for training-related effects (participants were not trained on this new melody). Before the study, we recorded a professional cellist playing the first five notes of a D major scale at 50 bpm on the MRI-compatible cello so that the recordings of the sequences could be used during training and scanning sessions. The recording was used as-is for the scale and was postprocessed using audio-editing software (Audacity) to reorder the pitches for the melodies, so that all stimuli had the same acoustical characteristics.

Training Sessions.
Playing task. Participants received individual cello lessons in a room at the McGill University faculty of music twice a week for 1 month. Lessons were administered by I.W. and M.S., in either French or English, adhering to a very precise protocol established for the study (see details in *SI Appendix, SI Methods*). The first two sessions lasted 1.5 h each; the next six sessions lasted 45 min each. During each lesson, participants practiced on the MRI-compatible cello while lying inside a cardboard structure that simulates the MRI scanner environment (*SI Appendix*, Fig. S1*D*). As in the scanner, soft foam cushions for head, arms, and legs were used to ensure that participants were comfortably positioned to avoid limb strain. Participants encountered and practiced both scales during session 1A and both melodies during session 1B. During the next six sessions, all four sequences were rehearsed. A metronome was used to help participants play at the tempo specified by the instructor. In all sessions, participants received feedback from the instructor on their pitch and tempo accuracy and sound quality.

The training protocol emerged and was formalized after pilot training testing. To this end, we produced a road map document to serve as a guideline for both instructors. Training covered a variety of exercises involving one hand only (e.g., bowing on open strings with right hand or practicing left finger placements) or both hands (e.g., playing along with the metronome at various tempi). During week 3 (sessions 4 and 5), emphasis was placed on tempo or pitch accuracy with exercises specifically focusing on rhythm (e.g., tapping the beat with the bow hand) or pitch (e.g., playing the first sequence interval along with the recording). During week 4 (sessions 6 and 7), participants practiced as in the scanner session; after listening to a musical sequence, they were asked to wait for two metronome beats before repeating the whole sequence on the cello with no metronome. Guidelines specifying the exact order of selected exercises per session, the number of times each exercise had to be performed, and the amount of time spent on each exercise were scrupulously respected. Therefore, this training procedure allowed very little scope for the use of different strategies during learning.

At the end of each training session the participant's performance was recorded twice at 50 bpm on each sequence. Participants were paced by an auditory metronome in these recordings. This information was used to determine performance accuracy and to quantify the participant's improvement over the course of training.

Imagery task. Participants also received training on imagery, with focus on the timing as well as vividness of mental imagery. For each sequence, they were asked to imagine playing the MRI-compatible cello in their mind without moving any articulators or producing any sounds. They were specifically instructed that they should feel the movement of specific articulators that would be associated with actual playing and hear their playing loud and clear in their mind (details are given in *SI Appendix, SI Methods*).

MRI Scan Sessions.
MRI data acquisition. Scanning sessions took place at the McConnell Brain Imaging Center at the Montreal Neurological Institute on a 3-T whole-body MR scanner (Siemens Trio) with a 32-channel head coil. Functional and structural MRI scans were collected at each of the three time points of the training. The protocol lasted 1 h in preSCAN, 1.5 h in SCAN wk1, and 2 h in SCAN wk4. All scan sessions included one sagittal T$_1$-weighted image (MPRAGE, voxel size 1 mm^3) for anatomical reference. We used a sparse-sampling paradigm for functional runs with a repetition time (TR) of 10 s. This paradigm minimizes the influence of the BOLD response due to scanner noise on the BOLD response to the task and minimizes the impact of playing-related movements on volume acquisition. It also allows auditory stimuli and playback responses to be heard in silence. One functional run contained 38 volumes in preSCAN and 98 volumes in both the other scan sessions. We recorded echo-planar imaging (EPI) images covering the whole head [voxel size 3.5 mm^3, 38 interleaved slices, echo time (TE) 30 ms, TR 10 s, flip angle = 90°]. One EPI ("dummy") volume was initially acquired and discarded to allow for T1-saturation effects. Resting-state scans were acquired at the start of the preSCAN and SCAN wk4. Additionally, diffusion-weighted images were acquired. Cello sounds were delivered via MRI-compatible headphones (S14; Sensimetrics Corp.) with foam inserts placed inside the ear canal. Stimuli were delivered diotically at a sound pressure level of 70 dB. Visual instructions were presented via back projection on a screen placed at the end of the MRI bore.

MRI protocol. During preSCAN, participants underwent four functional runs of listening-only trials. Each sequence was presented 24 times across four runs, which yielded 96 listening trials in total. Because participants at this point in the procedure could not yet play or imagine playing unfamiliar melodies on a new instrument, the playing and imagining tasks were not included in preSCAN. The same functional protocol was used during both SCAN wk1 and SCAN wk4. It comprised four identical functional runs. Each run was composed of nine blocks of trials. In the first six blocks, trials were grouped by three (dark blue in *SI Appendix*, Fig. S1*C*). Within each block, participants had to perform three tasks for a particular sequence: (*i*) listen, (*ii*) play, and (*iii*) play with no auditory feedback (hereafter called "playnoA"; details are given in *SI Appendix, SI Methods*). The listening task always appeared first, and the order of presentation of the playing and playnoA tasks for that particular sequence was counterbalanced across blocks. Participants did not know whether they would have the auditory feedback before starting to play and had not received any training on that task. Audio signals of the performance were systematically recorded. The order of the four sequences was counterbalanced across the four runs. After those six blocks, participants underwent a block of six listening trials of the untrained new melody (orange in *SI Appendix*, Fig. S1*C*). Finally, in the last two blocks, trials were grouped by four (light blue in *SI Appendix*, Fig. S1*C*). Within each block, participants had to imagine each of the four sequences. The order of the four sequences was randomized across subjects. All blocks were interspersed with trials of rest, with one rest trial at the end of every block and two rest trials at the end of every three blocks. Rest trials were used as the baseline control condition in the subsequent contrast analyses. This protocol yielded 96 trials in total for the listening, playing, and playnoA tasks (24 trials per sequence for each task) and 32 trials in total for the imagining task (eight trials per sequence). This paper focused on the listening and playing tasks.

On each trial, first a visual instruction for the task (e.g., "Listen") and the name of the sequence (e.g., "sequence A") was presented for 2.3 s, corresponding to the scanner acquisition time of the preceding trial. A series of two flashes with a constant interstimulus interval of 1.2 s. (i.e., at 50 bpm) accompanied all instructions. These flashes were used as visual pacing stimuli. Then, participants had to perform the 6-s task within the subsequent 7.7 s of silence. They were specifically instructed to begin the task on the expected third flash, following the temporal interval formed by the preceding flashes. Participants were given 5 min at the beginning of SCAN wk1 and SCAN wk4 to practice the sequences while they were lying on the scanner's bed but without scanning.

Behavioral data analysis. Performance was evaluated using two accuracy indexes calculating average pitch and rhythm deviations from the target sequences. All audio recordings were automatically analyzed using a MATLAB-written script developed for the purpose of this study. We used an ad-hoc onset-detection algorithm, operating in the time domain, that detects significant changes in the energy envelope of the signal. This method allows the

identification of the portion of the signal corresponding to each note in a sequence. We used the YIN algorithm to estimate pitch levels (65). Ultimately, the script compares the sequence of notes produced by the subject with the sequence template, matching the onsets of the first note of both played and target sequences, and calculates an average pitch deviation (in cents; referred to in the text as "pitch error") and a timing average deviation (in milliseconds; referred to in the text as "tempo error") from target melodies. These two behavioral indexes were analyzed by repeated-measures ANOVA.

Performance score. For each subject in a scan session, one mean pitch error and one mean tempo error were calculated by averaging pitch and tempo deviation indices, respectively, across all sequences played. Then, for each subject, a single performance score was derived from the mean pitch error and the mean tempo error to represent individual training achievements. The performance score was obtained by linear rescaling the mean pitch and tempo errors on an arbitrary 0–100 scale and by taking the arithmetic mean of ($-1 \times$ mean pitch error) and ($-1 \times$ mean tempo error). That way, the greater the score, the better was the performance inside the scanner.

fMRI data analysis. Functional data were processed and analyzed using the FSL software tool (FSL Technologies, Ltd.), except for the FC analyses (see below). Nonbrain voxels were removed using the Brain Extraction Tool (BET) of the FSL software. Images were motion-corrected using MCFLIRT of the FSL software, spatially smoothed using a Gaussian kernel (5 mm FWHM), and were high-pass filtered (100 Hz). Individual fMRI data were registered to the individual's T1-weighted anatomical images (three-parameter linear transformation) and registered to Montreal Neurological Institute (MNI) standard space for third-level analyses (12-parameter linear transformation). Because we used a sparse sampling design, we did not apply any slice-timing correction. Task-related BOLD signals from each voxel were modeled by using the general linear model (FEAT) of the FSL software, including in the model all conditions that were present during the scanning session.

Contrasts. For each individual scan, contrast images (task vs. rest) were computed to assess basic task-related activity and, in turn, changes in activity due to training. For all within-scan analyses, first-level contrast maps for each subject and run were entered into fixed-effects second-level analyses for each subject. Then third-level group analyses and between-scans analyses were made in a random effects model in MNI space using FLAME1 of the FSL software. Results were thresholded at z >2.3 and cluster-corrected P < 0.05, using a Z statistic threshold to define contiguous clusters, followed by estimation of the significance level of each cluster based on the cluster probability threshold. Anatomical localization was determined using the Juelich histological atlas and the Harvard–Oxford cortical and subcortical structural atlases, which are part of the FSL package.

Conjunctions. Conjunction analysis (66) was performed over two contrasts to determine the intersection of suprathreshold brain regions across those two statistic images. Results were thresholded at z >2.3, i.e., P < 0.05 uncorrected.

Regressions. To examine training-related brain changes as a function of performance on an individual basis, regression analyses were performed based on the performance score at SCAN wk4, which represents the individual's final performance level. To reveal markers of predisposition for cello training, regressions of pretraining activity using behavioral scores as regressor were assessed statistically at a more stringent height threshold (z >2.6, i.e., P < 0.01 uncorrected). Following the method described in ref. 35, we also employed a leave-one-participant-out cross-validation procedure in which a single subject is iteratively left out of the group regression analysis. The group analysis defines an ROI which is then applied to the data collected from the participant left out. Subsequently, activity in the functionally defined ROI is extracted for the participant left out and is used for the computation of the correlation. The procedure is then repeated for each participant. The regression analysis from the remaining participants

thus serves as an independent localizer for the participant left out. All ROIs were set to be similar in size to the one defined by the regression analysis taking all 13 participants. Post hoc functional ROI analyses were performed using Featquery in FSL and were tested for correlations between pretraining activations and behavioral achievements.

Functional connectivity analyses. We conducted task-related FC analyses to assess the network relationships between different regions during task performance. We also conducted resting-state FC analyses because connectivity during rest has been proposed to reflect intrinsic brain-connectivity networks (67). These networks have been shown to be shaped by repetitive long-term training in various domains such as visual learning (68), motor learning (69), or musical training (70) and may also serve as predictors of subsequent learning success (71). Both task-related FC and resting-state FC analyses were performed using the CONN-fMRI toolbox for SPM (https://www.nitrc.org/projects/conn), which computes the temporal correlation between the BOLD signals from given seed regions and other voxels in the brain. Data were preprocessed within the toolbox. For task-related FC, because we used a sparse sampling design, we did not apply any slice timing correction. For resting-state FC, preprocessing included the slice-timing correction for interleaved acquisitions. Data were bandpass filtered (0.008–0.09 Hz). Second-level analysis was performed using the general linear model within CONN to determine between-scan group differences in correlation maps. We performed seed-to-ROI analyses (i.e., seed-based approach with hypothesis-led target ROIs) and confirmed them with seed-to-voxel analyses (i.e., seed-based approach with all other voxels of the brain). The analyses compute the correlation between the mean time series within the seed and the average time series within each target ROI (seed-to-ROI analysis) or at each voxel (seed-to-voxel analysis). The results of seed-to-ROI FC analyses are presented in the main text, and FC maps resulting from whole-brain seed-to-voxel analyses are presented in *SI Appendix*. Seed-to-ROI FC maps were corrected for multiple comparisons by using false-discovery rate (FDR) correction thresholded at P < 0.05 (unless stated otherwise). Seed-to-voxel FC maps were cluster-corrected for multiple comparisons at P < 0.05 by using a P statistic threshold to define contiguous clusters showing positive correlation followed by estimation of the significance level of each cluster based on the cluster probability threshold.

ROI definition. Based on a priori-defined hypotheses, in the seed-to-ROI analyses we restricted the search volume to the bilateral AC, dPMC, and preSMA and SMA. ROIs in the AC [including right and left posterior STG (pSTG) and PT] and SMA (including the right and left SMA ROI) were used as provided by the toolbox from anatomical regions labeled by the FSL Harvard–Oxford atlas. For the dorsal premotor regions, because no differentiation is made between different subregions of the premotor area, we extracted voxel coordinates for peak activity in the dorsal region from the whole-brain group contrast of play vs. rest at SCAN wk4 and used them as the center coordinates for spherical ROIs (8-mm radius). Center coordinates are reported in MNI space for bilateral dPMC (left dPMC: 71 60 61; right dPMC: 18 62 61). The ROI encompassing the preSMA was defined functionally based on the results of the regression that tested for correlations between the pretraining BOLD response during listening and the performance score at SCAN wk4 (*Results*). This post hoc functional ROI was created from the resulting image thresholded at z >2.6.

1. Palmeri TJ, Wong AC, Gauthier I (2004) Computational approaches to the development of perceptual expertise. *Trends Cogn Sci* 8:378–386.
2. Zatorre RJ, Fields RD, Johansen-Berg H (2012) Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nat Neurosci* 15:528–536.
3. Janata P, Grafton ST (2003) Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nat Neurosci* 6:682–687.
4. Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: Auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8:547–558.
5. Herholz SC, Zatorre RJ (2012) Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76:486–502.
6. Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11:599–605.
7. Wan CY, Schlaug G (2010) Music making as a tool for promoting brain plasticity across the life span. *Neuroscientist* 16:566–577.
8. Chen JL, Rae C, Watkins KE (2012) Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. *Neuroimage* 59:1200–1208.
9. 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. *Eur J Neurosci* 24:955–958.
10. Herholz SC, Coffey EB, Pantev C, Zatorre RJ (2016) Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cereb Cortex* 26:3125–3134.
11. Lahav A, Saltzman E, Schlaug G (2007) Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *J Neurosci* 27:308–314.
12. Lappe C, Herholz SC, Trainor LJ, Pantev C (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. *J Neurosci* 28:9632–9639.
13. Bangert M, Altenmüller EO (2003) Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neurosci* 4:26.
14. Bangert M, et al. (2006) Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage* 30:917–926.
15. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.

NEUROSCIENCE

16. Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensori-motor control and integration. *Hear Res* 271:16–25.
17. Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex* 18:2844–2854.
18. Wollman I, Fritz C, Poitevineau J, McAdams S (2014) Investigating the role of auditory and tactile modalities in violin quality evaluation. *PLoS One* 9:e112552.
19. Wollman I, Fritz C, Poitevineau J (2014) Influence of vibrotactile feedback on some perceptual features of violins. *J Acoust Soc Am* 136:910–921.
20. Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1995) Increased cortical representation of the fingers of the left hand in string players. *Science* 270:305–307.
21. Bangert M, Schlaug G (2006) Specialization of the specialized in features of external human brain morphology. *Eur J Neurosci* 24:1832–1834.
22. Behroozmand R, et al. (2016) Neural correlates of vocal production and motor control in human Heschl's Gyrus. *J Neurosci* 36:2302–2315.
23. Burnett TA, Larson CR (2002) Early pitch-shift response is active in both steady and dynamic voice pitch control. *J Acoust Soc Am* 112:1058–1063.
24. Zarate JM, Zatorre RJ (2008) Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40:1871–1887.
25. Zatorre RJ (2013) Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science* 342:585–589.
26. Schneider P, et al. (2002) Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat Neurosci* 5:688–694.
27. Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 23:5545–5552.
28. Menning H, Roberts LE, Pantev C (2000) Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport* 11:817–822.
29. Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009) Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex* 19:1583–1596.
30. Gaser C, Schlaug G (2003) Brain structures differ between musicians and non-musicians. *J Neurosci* 23:9240–9245.
31. Jäncke L, Gaab N, Wüstenberg T, Scheich H, Heinze HJ (2001) Short-term functional plasticity in the human auditory cortex: An fMRI study. *Brain Res Cogn Brain Res* 12:479–485.
32. Zatorre RJ, Delhommeau K, Zarate JM (2012) Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Front Psychol* 3:544.
33. Hollinger AD, Wanderley MM (2013) MRI-compatible optically-sensed cello. *SENSORS, 2013* (IEEE, Piscataway, NJ), pp 1–4.
34. Pernet CR, Wilcox R, Rousselet GA (2013) Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Front Psychol* 3:606.
35. Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S (2010) Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage* 50:572–576.
36. Pfordresher PQ, Mantell JT, Brown S, Zivadinov R, Cox JL (2014) Brain responses to altered auditory feedback during musical keyboard production: An fMRI study. *Brain Res* 1556:28–37.
37. Kleber B, Veit R, Birbaumer N, Gruzelier J, Lotze M (2010) The brain of opera singers: Experience-dependent changes in functional activation. *Cereb Cortex* 20:1144–1152.
38. Segado M, Hollinger A, Thibodeau J, Penhune V, Zatorre RJ (2008) Partially overlapping brain networks for singing and cello playing. *Front Neurosci* 12:351.
39. Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: The action simulation for auditory prediction (ASAP) hypothesis. *Front Syst Neurosci* 8:57.
40. Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. *Proc Natl Acad Sci USA* 114:E8913–E8921.
41. Keller PE (2012) Mental imagery in music performance: Underlying mechanisms and potential benefits. *Ann N Y Acad Sci* 1252:206–213.
42. Stephane MA, Lega C, Penhune V (2018) Auditory prediction cues motor preparation in the absence of movement. *Neuroimage* 174:288–296.
43. Brown RM, et al. (2013) Repetition suppression in auditory-motor regions to pitch and temporal structure in music. *J Cogn Neurosci* 25:313–328.
44. Goodale MA (2011) Transforming vision into action. *Vision Res* 51:1567–1587.
45. Singh P, Jana S, Ghosal A, Murthy A (2016) Exploration of joint redundancy but not task space variability facilitates supervised motor learning. *Proc Natl Acad Sci USA* 113:14414–14419.
46. Stephan MA, Heckel B, Song S, Cohen LG (2015) Crossmodal encoding of motor sequence memories. *Psychol Res* 79:318–326.
47. Stephan MA, Brown R, Lega C, Penhune V (2016) Melodic priming of motor sequence performance: The role of the dorsal premotor cortex. *Front Neurosci* 10:210.
48. Wong PC, Perrachione TK, Parrish TB (2007) Neural characteristics of successful and less successful speech and word learning in adults. *Hum Brain Mapp* 28:995–1006.
49. Watanabe T, Yagishita S, Kikyo H (2008) Memory of music: Roles of right hippocampus and left inferior frontal gyrus. *Neuroimage* 39:483–491.
50. Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6:37–46.
51. Grahn JA, Rowe JB (2013) Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cereb Cortex* 23:913–921.
52. Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015) Finding the beat: A neural perspective across humans and non-human primates. *Philos Trans R Soc Lond B Biol Sci* 370:20140093.
53. Albouy P, et al. (2013) Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain* 136:1639–1661.
54. Grimault S, et al. (2014) Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *Neuroimage* 94:96–106.
55. Doyon J, Benali H (2005) Reorganization and plasticity in the adult brain during learning of motor skills. *Curr Opin Neurobiol* 15:161–167.
56. Lima CF, Krishnan S, Scott SK (2016) Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends Neurosci* 39:527–542.
57. Jueptner M, et al. (1997) Anatomy of motor learning. I. Frontal cortex and attention to action. *J Neurophysiol* 77:1313–1324.
58. Tanji J (2001) Sequential organization of multiple movements: Involvement of cortical motor areas. *Annu Rev Neurosci* 24:631–651.
59. Dayan E, Cohen LG (2011) Neuroplasticity subserving motor skill learning. *Neuron* 72:443–454.
60. Lehéricy S, et al. (2005) Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proc Natl Acad Sci USA* 102:12566–12571.
61. Woods EA, Hernandez AE, Wagner VE, Beilock SL (2014) Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain Cogn* 87:122–133.
62. Margulis EH, Mlsna LM, Uppunda AK, Parrish TB, Wong PC (2009) Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Hum Brain Mapp* 30:267–275.
63. Grahn JA, McAuley JD (2009) Neural bases of individual differences in beat perception. *Neuroimage* 47:1894–1903.
64. Peretz I, Champod AS, Hyde K (2003) Varieties of musical disorders. *Ann N Y Acad Sci* 999:58–75.
65. de Cheveigné A, Kawahara H (2002) YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am* 111:1917–1930.
66. Price CJ, Friston KJ (1997) Cognitive conjunction: A new approach to brain activation experiments. *Neuroimage* 5:261–270.
67. Raichle ME (2010) The brain's dark energy. *Sci Am* 302:44–49.
68. Urner M, Schwarzkopf DS, Friston K, Rees G (2013) Early visual learning induces long-lasting connectivity changes during rest in the human brain. *Neuroimage* 77:148–156.
69. Vahdat S, Darainy M, Milner TE, Ostry DJ (2011) Functionally specific changes in resting-state sensorimotor networks after motor learning. *J Neurosci* 31:16907–16915.
70. Klein C, Liem F, Hänggi J, Elmer S, Jäncke L (2016) The "silent" imprint of musical training. *Hum Brain Mapp* 37:536–546.
71. Ventura-Campos N, et al. (2013) Spontaneous brain activity predicts learning ability of foreign sounds. *J Neurosci* 33:9295–9305.