2021

Mapping Specific Mental Content During Musical Imagery

Mor Regev
Montreal Neurological Institute, mor.regev@mcgill.ca

Andrea R. Halpern
Bucknell University, ahalpern@bucknell.edu

Adrian Owen
Western University, aowen6@uwo.ca

Aniruddh Patel
Tufts University, a.patel@tufts.edu

Robert J. Zatorre
McGill University

Cerebral Cortex, 2021;00: 1–19

doi: 10.1093/cercor/bhab036
Original Article

ORIGINAL ARTICLE

Mapping Specific Mental Content during Musical Imagery

Mor Regev[1,2,3], Andrea R. Halpern[4], Adrian M. Owen[5,6], Aniruddh D. Patel[6,7] and Robert J. Zatorre[1,2,3,6]

[1]Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada, [2]International Laboratory for Brain, Music and Sound Research, Montreal, QC H2V 2J2, Canada, [3]Centre for Research in Language, Brain, and Music, Montreal, QC H3A 1E3, Canada, [4]Department of Psychology, Bucknell University, Lewisburg, PA 17837, USA, [5]Brain and Mind Institute, Department of Psychology and Department of Physiology and Pharmacology, Western University, London, ON N6A 5B7, Canada, [6]Canadian Institute for Advanced Research, Brain, Mind, and Consciousness program and [7]Department of Psychology, Tufts University, Medford, MA 02155, USA

Address correspondence to Mor Regev. Email: mor.regev@mcgill.ca.

Abstract

Humans can mentally represent auditory information without an external stimulus, but the specificity of these internal representations remains unclear. Here, we asked how similar the temporally unfolding neural representations of imagined music are compared to those during the original perceived experience. We also tested whether rhythmic motion can influence the neural representation of music during imagery as during perception. Participants first memorized six 1-min-long instrumental musical pieces with high accuracy. Functional MRI data were collected during: 1) silent imagery of melodies to the beat of a visual metronome; 2) same but while tapping to the beat; and 3) passive listening. During imagery, inter-subject correlation analysis showed that melody-specific temporal response patterns were reinstated in right associative auditory cortices. When tapping accompanied imagery, the melody-specific neural patterns were reinstated in more extensive temporal-lobe regions bilaterally. These results indicate that the specific contents of conscious experience are encoded similarly during imagery and perception in the dynamic activity of auditory cortices. Furthermore, rhythmic motion can enhance the reinstatement of neural patterns associated with the experience of complex sounds, in keeping with models of motor to sensory influences in auditory processing.

Key words: auditory imagery, fMRI, inter-subject correlation, music, rhythmic motion

Introduction

Mental imagery, the conscious representation of sensory information without direct external stimulus, is arguably one of the primary components of human cognition. It plays an important role in memory retrieval, future planning, decision making, creativity, and emotional regulation (Baddeley and Logie 1992; Moulton and Kosslyn 2009; Lucas et al. 2010; Keogh and Pearson 2011; Palmiero et al. 2011; Amit and Greene 2012; Schacter et al. 2012). In mental imagery, previously experienced sensory content occupies our conscious mind, reinstating old events or composing episodes based on novel sensory combinations.

Despite its great allure, the subjective nature of this covert experience and the lack of clear observable behavioral markers make its experimental investigation notoriously difficult (Hubbard 2018). Thus, despite progress in showing similar global neural responses to perceived and imagined events, we do not really understand how the actual contents of internal sensory experiences are represented in the brain, particularly in the

case of auditory imagery, which has been less studied than visual imagery. Are the neural codes elicited by perception and imagery of the same auditory event similar? Do they represent the temporally unfolding auditory content in our internal thought?

At a phenomenological level, many sensory and cognitive aspects of perceived sounds are reported to also be experienced in imagery, and imagery of music is no exception (Lucas et al. 2010, see Hubbard 2010 for a review). Yet, the neural realization of these similarities is not well specified. We know that the general topography of brain areas recruited when people imagine and perceive musical sounds is similar. Bilateral auditory cortices are consistently recruited in both imagery and perception of sounds (Zatorre and Halpern 1993; Griffiths 2000; Halpern 2001; Halpern et al. 2004; Kraemer et al. 2005; Herholz et al. 2012; Ding et al. 2019), as are frontal and parietal cortical structures such as the supplementary motor area (Zatorre et al. 1996; Halpern and Zatorre 1999; Hickok et al. 2003; Foster and Zatorre 2009; Herholz et al. 2012; Foster et al. 2013; Ding et al. 2019). Thus, prior studies have clearly established that individuals are able to voluntarily reinstate musical content in their "mind's ear", and that there is a shared neural substrate for perception and imagery. However, it is still a mystery whether these neural substrates represent the *specific content* of experience, regenerated during imagery and evolving with the flow of the conscious mind (James 1890), or rather only share a global involvement in both imagery and perception.

The neural representation of internal reinstatement of complex sounds has been most commonly studied by looking for spatially overlapping neural responses while listening versus imagining sounds. Yet, a brain area could respond reliably to both internal and external sounds without the same operations necessarily being carried out in each case, and with the same temporal dynamics (Dinstein et al. 2007; Ben-Yakov et al. 2012; Regev et al. 2019). Such findings are therefore limited in terms of differentiating between representations of specific imagined content reinstated from memory (e.g., one sound vs. another).

More recently, studies have investigated the different information that might be represented within a given region of interest during auditory imagery, using decoding methods. Most of these efforts focused on spoken language, making great progress in localization of the content-specific representation of imagined linguistic constructs, ranging from single vowels to full sentences (e.g., Pei et al. 2011; Ikeda et al. 2014; Martin et al. 2014; Rampinini et al. 2017; Cervantes Constantino and Simon 2018; Musch et al. 2020). Nevertheless, the dominant semantic context and linguistic knowledge relevant for these paradigms can strongly influence the recruited sensory circuits that are involved in auditory imagery (Kraemer et al. 2005). It could be argued that when using internal speech, the identified neural representations, especially outside early auditory cortices, might in fact capture more abstract or symbolic processes unique to language, rather than reinstatement of auditory experience per se.

Other studies used imagined environmental sounds to identify spatially distributed neural representations, mostly in associative auditory cortices (Meyer et al. 2010; Vetter et al. 2014; Linke and Cusack 2015; also see de Borst et al. 2016 for findings in somatomotor areas). While the ability to decode which environmental sound a person is imagining from fine-grained spatial patterns of cortical activity does not involve linguistic mechanisms present during internal speech, such findings do not address two critical issues. First, it is not clear if the decoded neural representations are the same during imagery as during listening to sounds, since direct evidence for reinstatement (during imagery) of neural patterns characteristic of auditory perception has been missing. Indeed, in contrast to findings in the visual system (Thirion et al. 2006; Albers et al. 2013), attempts to identify any reinstated neural patterns in auditory brain regions during imagery have either failed, or have observed them only in the somatomotor cortex (Vetter et al. 2014; Gu et al. 2019). Second, environmental sounds are mostly brief by nature and are thus missing the temporal dynamics that occur over an extended period of time. Given that humans rely heavily on complex auditory sequences for communication via speech and music, neural studies of environmental sounds cannot fully capture the temporal complexity and structural dependencies of auditory content crucial to human cognition. It therefore remains important to test whether the same neural populations share the representation of sound sequences across perception and imagery, and whether the temporal pattern of activity in these populations is associated with the specific nonlinguistic content of complex sensory sequences.

Similarities in the representation of sounds across imagery and perception can be studied not only locally, but also by comparing interactions across neural populations. One of the most prominent neural attributes of complex sound perception is the interaction between the auditory and motor networks (Zatorre et al. 2007; Hickok et al. 2011; Morillon et al. 2015). The effect of movement on auditory perception and its related neural activity has been demonstrated mostly in the context of the modulation of self-produced sounds, such as speech or music, by associated movements such as articulation or musical performance (Zatorre et al. 2007; Rauschecker and Scott 2009; Hickok et al. 2011; Repp and Su 2013; Rauschecker 2018; Reznik and Mukamel 2019).

Recent studies have shown that rhythmic tapping to ongoing external auditory stimuli can also modulate their perception by facilitating selection of relevant auditory events and refining the representation of pitch (Morillon et al. 2014; Nozaradan et al. 2016; Morillon and Baillet 2017). These results suggest that even when movements are not the source of the perceived sounds, their co-occurrence is sufficient to induce modulation in the auditory response. Interestingly, there is evidence showing that rhythmic movement also affects the mnemonic representation of sounds. For example, tapping along during an initial exposure to music has been shown to increase its memorability and chances of becoming an earworm (Mikumo 1994; McCullough Campbell and Margulis 2015). These results, together with the evidence for the neural coactivation of motor regions and sensory-motor interaction during auditory imagery (Li et al. 2020, also see Lima et al. 2016 for a review), raise the possibility that the strong auditory–motor relationship might allow movement to modulate the internal representation of sounds.

In this work, we compared the unique temporal neural response profile of imagined and heard musical pieces. This approach allowed us to probe the specific contents of imagined auditory experiences that unfold over time, and also to explore the influence of rhythmic motion on the neural representation of auditory imagery. Prior to scanning, participants memorized six different ~1-min-long instrumental pieces of real music, and the accuracy of their musical imagery was verified behaviorally. Next, brain activity was measured with functional magnetic resonance imaging (fMRI) as participants either listened to each of the memorized melodies, or imagined each melody under two

conditions: with and without simultaneous rhythmic tapping to a visual metronome. In this manner, the timing of the auditory stimuli and imagery were precisely matched across conditions and participants, allowing us to measure temporal alignment of brain responses.

We used inter-subject correlation (ISC) analyses, which provide a tool for extracting neural activity that is locked in time to continuous natural stimuli (Hasson et al. 2004; Hasson et al. 2009; Simony et al. 2016). By measuring the reliability of neural response time courses across brains, we can detect neural activity that is specifically induced by temporally extended experiences shared by participants. For example, previous studies using this method have shown that listening to the same music (Alluri et al. 2012; Abrams et al. 2013; Farbood et al. 2015) or vocalizing the same learned story (Silbert et al. 2014) can induce strong between-participant similarity in the time courses of brain activity in many regions. We extend this approach to investigate neural commonalities between people during purely *internally driven* mental processes, covertly generated with no external manifestation. ISC analyses also allow us to measure ongoing components that are shared across different experimental conditions, akin to the way the neural response to a narrative has been compared across different sensory forms of presentation (Honey et al. 2012; Regev et al. 2013; Nguyen et al. 2019). Here, we used this method to ask whether the shared content of inner conscious experience elicits a similar neural code across people, and whether this code is common to the external perceptual experience as well. Such a finding could demonstrate how the subjective experience of our inner and outer landscape is largely based on a common set of principles.

Thus, we first tested the prediction that imagery of auditory content can evoke similar response patterns across people, despite the lack of external input. Next, by comparing temporal response patterns across perception and imagery, we investigated whether responses during perception were reinstated during imagery in a melody-specific manner, allowing us to test the hypothesis that the specific contents of the conscious experience could be mapped. Finally, we studied the effect of rhythmic tapping on the reinstatement of melody-specific activity, to test the idea that top-down motor signals enhance internal auditory representations.

Materials and Methods

Participants

A total of 25 participants were included in the final analysis (22 females, 3 males; age 18–32, mean age 21.9). Out of the initial 44 participants recruited, 17 were discarded for failing to memorize the melodies sufficiently, as assessed by the external-recall test (see *Learning phase*), and a further two were discarded from the analysis: one due to corrupted functional data and one due to head motion greater than 2 mm.

Participants were selected without regard to their musical background (see Supplementary Methods), but were required to have learned at least three melodies and recalled them in the past by either vocalizing or playing an instrument. Their self-reported auditory imagery ability, as measured using the Bucknell Auditory Imagery Scale (BAIS), was not different than what has been previously reported in a larger sample of college students (current sample: $M = 5.4$ $SD = 0.7$; previous sample: $N = 76$, $M = 5.1$ $SD = 0.9$; unpaired two-tailed test: $t_{(99)} = 1.52$, $p = 0.12$; Halpern 2015). All participants were right-handed, reported normal hearing and no absolute pitch, and did not report any neurological disorders. Procedures were approved by the Research Ethics Board of the Montreal Neurological Institute.

MRI Acquisition

Participants were scanned in a 3 T full-body MRI scanner (Prisma, Siemens) with a 64-chanels head coil. Functional images were obtained with an interleaved multiband echo planer imaging (EPI) sequence (TE = 39 ms; flip angle = 50°; multislice factor = 4; field of view [FOV] = 192 ×192 mm^2; echo spacing = 0.55 ms; 72 oblique axial slices), resulting in a voxel size of 2.0 mm isotropic and a TR of 1500 ms. Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence (TR = 2300 ms; TE = 2.98 ms; flip angle = 9°; 1.0 mm^3 resolution; FOV = 256 mm^2).

Stimuli

The auditory stimulus consisted of six distinct instrumental melodies (with no vocals) composed by Joe Hisaishi for three different animated movies soundtracks: “Castle in the Sky” (melodies [1] “Discouraged Pazu” and [2] “Carrying You”), My Neighbor Totoro (melodies [3] “Evening Wind” and [4] “Let’s Go to the Hospital”) and “Spirited Away” (melodies [5] “One Summer’s Day” and [6] “Reprise”). These melodies were written to evoke an emotional response in the audience, and exhibited full musical complexity, containing multiple instruments and melodic lines. At the same time, they also have a simple and catchy lead melodic line, making their memorization relatively easy. From each original track, we identified about a minute-long segment of the music ([1] 46 s [2] 74 [3] 78 [4] 84 [5] 91 [6] 44), which contains clear and dynamic melodic line and a stable tempo.

The six melodies were split into two subgroups, and the tempo of each three melodies was matched. To match the tempo, we first detected the beat locations for each original melody using the DBNBeatTracker from madmom library (Böck et al. 2016). Next, each melody was adjusted so that the average number of beats per minute matched the subgroup average (73 BPM for melodies 1, 3, & 5; 107 BPM for melodies 2, 4, & 6) and the beats were aligned (using the librosa library https://github.com/librosa/librosa/releases/tag/0.5.1). Thus, the melodies within each tempo subgroup shared an identical timing of beats. In addition, the waveforms of all melodies were matched for an averaged Root-Mean-Square value (implemented using librosa; the melodies can be found here: https://doi.org/10.5281/zenodo.3993675).

Experimental Design

Learning Phase

Participants were asked to memorize the six melodies at home, during a period of 1–2 weeks, and spread their learning across at least 2 days. In addition to the melodies, they were provided with home-test videos that assisted in practicing accurate out loud recall of the melodies, in preparation for the memory test to come. These videos began with the first 6 s of the music, accompanied by a synchronized visual metronome in the shape of a white bouncing ball at the center of the black screen. After the music had stopped, the metronome carried on in silence, allowing participants to hum the rest of the melody to the

beat. The music returned for the melody's last 3 s, and thus provided participants with feedback whether their humming was correctly aligned with the original melody. The videos were accessed by participants through an online video player (Wistia.com), which enabled the experimenters to track the time participants spent listening to and recalling the melodies. As a condition for attempting the memory test and prove their knowledge, participants were asked to listen to each of the melodies at least three times, and use the home-test video at least three times successfully. Other than that, participants were not given any instructions regarding the strategy they should take for learning.

At the end of the learning period, a humming test was performed in the lab to assess participants' ability to accurately recall the melodies. In this test, participants were shown the same video as used earlier during recall practice described above and were asked to hum the melodies out loud. A trained experimenter assessed the accuracy of their recall by evaluating the melodic contour and the temporal precision. Voicing quality, such as pitch accuracy or voice stability, was not considered, in an attempt to dissociate participant's knowledge of the melodies from general singing abilities. To pass the test, participants had to fully recall each of the six melodies to the beat of the metronome, without skipping any beat. Successful participants were included in the experiment and continued to perform the internal memory test (see *Behavioral assessment of imagery*) and the fMRI scan.

Behavioral Assessment of Imagery

Before the scanning session, we assessed participants' ability to accurately recall the melodies internally, without overt vocalization or lip movement. Each participant performed two memory tests (adopted from Herholz et al. 2008; Weir et al. 2015): (1) A "Beat" test, designed to capture accuracy of the recall of the melody in time, keeping the rhythm of the presented metronome without skipping a beat and (2) a "Pitch" test, designed to capture the ability to keep track of the original pitch of the internally recalled melody. Each test included three rounds in random order; each one assessed a different pair of melodies: [1] + [6], [2] + [3], or [4] + [5]. Each test was performed under two movement instructions: during melody recall, participants were either asked to tap their right finger to the rhythm of a visual metronome or to stay completely still (considering any body movement, such of the finger, mouth, head, or limbs).

Beat Test. Each trial in this test began with the first 6 s of the music, accompanied by a synchronized visual metronome in the shape of a white bouncing ball at the center of the black screen. After the music had stopped, the metronome carried on in silence, allowing participants to imagine the rest of the melody to the beat. The music had returned for the last few seconds of the melody, and participants were then asked to report whether that last musical segment (probe) was correctly placed in time, considering the prior period of imagery (possible responses: yes; maybe yes; maybe no; no). Whereas half of these musical probes played the accurate continuation of the melody, the other half played an incorrect segment of the music (two beats early or late in the musical stream). For each melody, there were two types of probe start time: one probe was designed to start playing on a beat which was the closest to the 6 s mark from the end of the melody *and* was the first beat in a bar ("first"). The second probe started two beats earlier than the other probe, hence around the middle of the previous bar ("middle"). Participants were tested on each melody eight times overall: while tapping – "first" played at the correct time, "first" played at an incorrect time, "middle" played at the correct time, "middle" played at an incorrect time; these were repeated without tapping. The incorrect probes played the music starting +/− 2 beats from the correct start time. Each test round of a pair of melodies included two blocks for each of the movement instructions (i.e., w/ tap or w/o tap). Each block contained four trials. The order of the two melodies and the type of probe presented were randomized within each test round.

Pitch Test. Each trial in this test began with the first 4 s of the music, accompanied by the visual metronome. After the music stopped, the metronome carried on in silence, allowing participants to imagine the rest of the melody to the beat for about 4 or 8 s. Next, when the music had returned for the next few seconds of the melody, participants were asked to report whether the musical probe was played in the correct pitch, considering the prior period of imagery. The probes always played the appropriate part of the melody time-wise, but although half of them were played in the original pitch of the melody, the other half was shifted by a major third up or down. Pitch modification were implemented using librosa and preserved the melodic contour of the sequence. For each melody, there were eight different trial start times that spread along the melody. The imagery lasted 4 s for half of the trials ("short") and 8 s for the other half ("long"). The imagery end times (which are also a probe's start times) were also spread along the melody. Participants were tested on each melody 64 times overall: while tapping – four different "short", each one twice in a correct pitch and twice in incorrect, four different "long", each one twice in a correct pitch and twice in incorrect; these were repeated without tapping. Each test round of a pair of melodies included four blocks for each of the movement instructions (i.e., w/ tap or w/o tap). Each block contained eight trials. The order of the two melodies and the type of probe presented was randomized within each test round. One participant did not perform this test due to equipment malfunction.

For both Beat and Pitch tests, the order of the rounds (three melody-pairs) was randomized for each participant. After each trial, feedback was provided based on performance, in the shape of a green (correct) or red (incorrect) dot at the center of the screen. The blocks of the two movement instructions were interleaved, and half of the participants started their test with a block with instruction to tap. During the imagery periods, background noises recorded from an EPI sequence were played (in SNR of 0.25; implemented using librosa) to simulate the experimental environment in the scanner and to prevent participants from hearing other noises such as their tapping sounds and heartbeats. The auditory stimulus was played to participants using sound-proof headphones (Vic Firth Stereo Isolation Headphones). Responses were recorded using a portable silicon keyboard mounted on an Ester foam, in order to minimize potentially distracting tapping noises. The tests were run on Psychophysics toolbox (Brainard 1997; Pelli 1997; Kleiner et al. 2007) for MATLAB (MathWorks) in a sound-proof room. Using a microphone and an inspection window, an experimenter monitored for any potential sounds and movements produced by the participants.

Right-tailed t tests ($\alpha = 0.05$) were conducted against chance level to assess the effect of memorization on accuracy of internal recall of melodies. Paired two-tailed t tests were conducted to compare the effect of tapping on accuracy of internal recall of melodies. Effect sizes were assessed using Cohen's d.

Scanning Procedure

In the scanner, each participant performed the following conditions for each of the six melodies: (1) passively listened to the melody ("perception"); (2) silently imagined the music to the rhythm of a visual metronome – a bouncing ball at the center of the screen – while keeping motionless ("imagery w/o tap"); and (3) silently imagined the music to the rhythm of a visual metronome while tapping the right index finger to the beat of the music ("imagery w/ tap"). The order of the melodies was randomized across participants, and each melody run started with the perception condition, followed by the two types of imagery conditions, in a random order (Fig. 1).

In the imagery conditions, participants were instructed to "imagine the melody to the rhythm of the bouncing ball". Each imagery condition began with the first 6 s of the music (rounded to the nearest beat), accompanied by the same visual metronome as in the memory test (see *Learning phase*). After the music stopped, the metronome carried on in silence, allowing participants to imagine the rest of the melody to the beat. After each imagery task, participants were asked to report their level of confidence in the accuracy of their imagery and the level of the experienced vividness, using 1 to 7 scales. Runs in which participants reported confidence in their accuracy equal or lower than 5 were repeated as many times needed, until confidence level was improved. For each participant, the final analyses included only a single imagery attempt for each melody in each condition which crossed this self-evaluation threshold (see more details in Supplementary Methods). The reported level of vividness was overall high across melodies and participants (w/o tap: M = 5.9, SD = 0.92; w/ tap: M = 6 SD = 1.06).

Participants also took part in control conditions, in which they watched the same visual metronome as in the imagery conditions while tapping to its beat ("control w/ tap") or keeping motionless ("control w/o tap"), but were not asked to imagine the music. These two control conditions repeated twice: once in 73 BMP and another time in 107 BPM, controlling for the two tempo subgroups of the melodies. The order of the control conditions was randomized across participants, with one BPM version at the beginning of the scan (i.e., before the perception and imagery conditions) and the other BPM version at the end of the scan.

Tapping and responses to questions were recorded using an MRI-compatible two-button box in the right hand. Participants were provided with an MRI-compatible in-ear mono earbuds (Sensimetrics model S14), which provided the same audio input in each ear. The volume of the auditory stimuli was adjusted individually for each participant to comfortable and clear level, and all the melodies were presented at the same volume level. Visual stimuli (metronome and self-report questions) were projected onto a rear-projection screen located behind the magnet bore, and were viewed with an angled mirror. Visual stimuli were created using the Psychophysics toolbox (Brainard 1997; Pelli 1997; Kleiner et al. 2007) and were combined with auditory stimuli using FFmpeg (http://www.ffmpeg.org/, version 4.0.2). The resulting audiovisual stimuli were presented and synchronized with MRI data acquisition onset using the Psychophysics toolbox.

During all runs, participants' eyes were observed by an experimenter (using Eyelink 1000 eye tracker) to ensure they were engaged with the task and not falling asleep. In addition, the experimenter listened to the sounds coming from an MRI-compatible microphone suspended approximately 1 in from participant's mouth, to ensure they were not vocalizing. Respiration was monitored during all runs at a sampling rate of 400 Hz using a belt fastened around the chest. The signal was transmitted from the scanner bed using a Seimens wireless device (respiratory sensor PERU).

Data Analysis

Preprocessing

The fMRI data were preprocessed in FSL (https://fsl.fmrib.ox.ac.uk), including slice time correction, motion correction, linear detrending, high-pass filtering (42 s period, corresponds to the duration of the shortest melody), spatial smoothing (6 mm FWHM Gaussian kernel), and coregistration and affine transformation of the functional volumes to a template brain (MNI152). As a first step before preprocessing of the data, we cropped out the signals of the first seven to eight TRs of the beginning of each melody. This included the removal the period in which the music was playing at the beginning of the imagery conditions (four or five TRs, depends on the tempo of the beat) as well as additional three TRs due to global arousal response after the end of the sound that added noise to the signal. All calculations were performed in volume space. Projections onto cortical surface for visualization were performed, as a final step, with NeuroElf (http://neuroelf.net). For voxel-wise analyses, functional images were resampled to 3 mm isotropic voxels.

Removal of Respiratory Signal Sources

Slow changes in respiration over time have been shown to induce robust changes in the BOLD signal (Chang et al. 2009) in many cortical regions, especially in multiband sequences (Scheel et al. 2014; Golestani et al. 2018). Therefore, we used multiple linear regression to project out from the BOLD data the respiratory variation per time (RVT) nuisance variable and the RETROICOR correction algorithm. These calculations were modified to account for the interleaved slice order with simultaneous acquired slices (Scheel et al. 2014). The respiratory variables were projected out separately for each participant (except two who were missing a respiratory signal due to equipment malfunction) for all conditions before performing the preprocessing.

ISC Analysis

We computed ISC in each condition for each of 400 parcels from independent whole-brain resting-state parcellation (Schaefer et al. 2018). ISC analysis maps were produced within each condition (i.e., motionless imagery, imagery while tapping, perception) of each melody, as well as between conditions (e.g., perception vs. imagery) of the same melody and across different melodies. The ISC maps provide a measurement of the reliability of brain responses to each of the melodies in each of the conditions by quantifying the correlation of the time course of the BOLD activity across participants who share the same experience (Hasson et al. 2004; Hasson et al. 2009). For each parcel, ISC within a condition is calculated as an average correlation $R = \frac{1}{N}\sum_{j=1}^{N} r_j$, where the individual r_j are the Pearson correlations between that parcel's BOLD time course in one individual and the average of that parcel's BOLD time courses in the remaining individuals.

ISC across conditions is calculated as an average $\tilde{R} = \frac{1}{N}\sum_{j=1}^{N} \tilde{r}_j$ over the correlations, $\tilde{r}_j$ between the BOLD time courses of the jth individual from the first condition and the average BOLD time courses of all individuals in the other condition. When the comparison was performed across different melodies, the time courses of the longer melody of the two were cropped to fit the duration of the other melody, before the ISC was calculated.

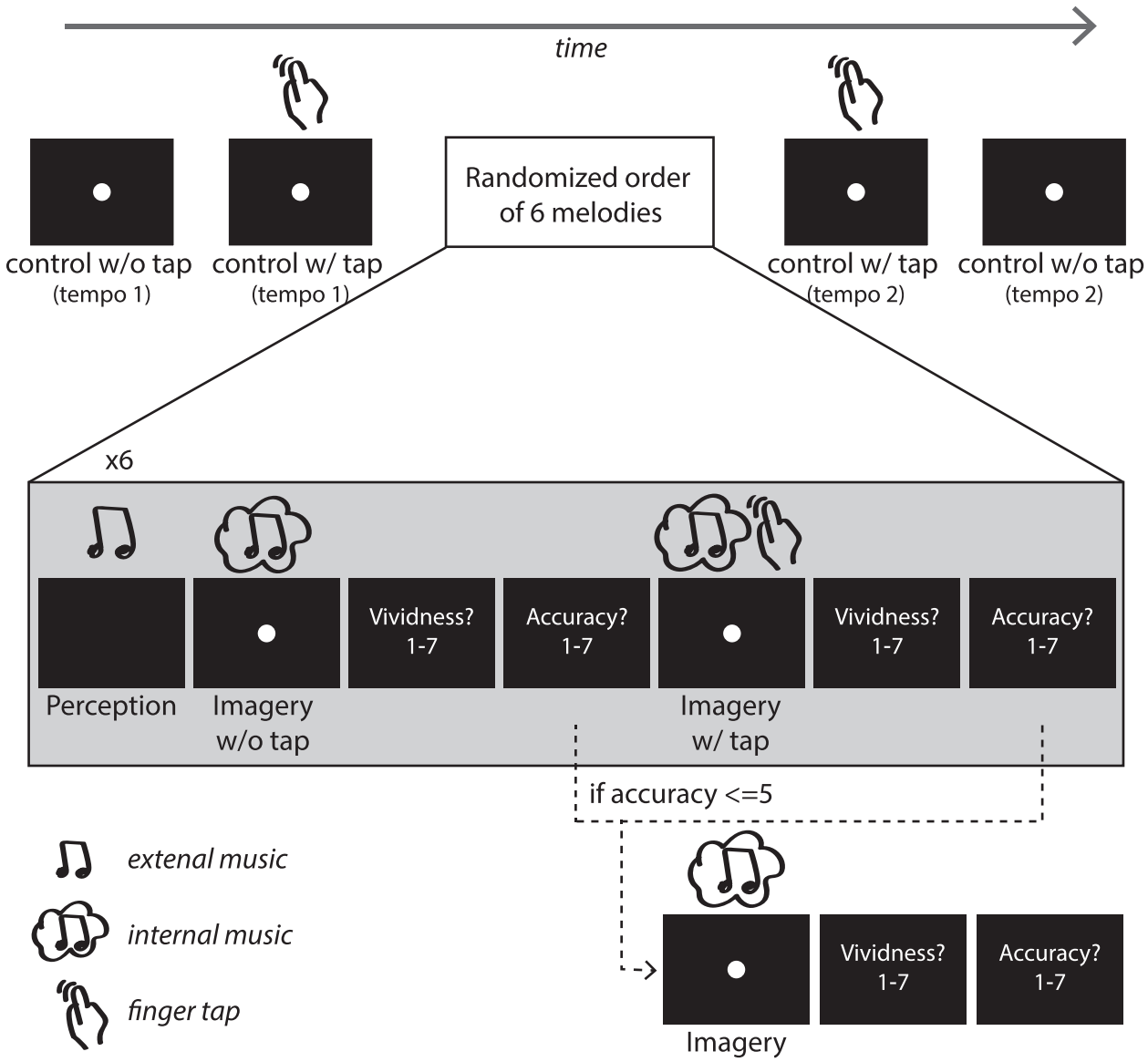

Figure 1. Schematic representation of the experimental design of an fMRI session. For each of the six melodies, participants passively listened to the melody ("perception") and then imagined it twice to the rhythm of a visual metronome (a bouncing white ball): while keeping motionless ("imagery w/o tap") and while tapping to the beat of the music ("imagery w/ tap"). After each imagery task, the level of confidence in accuracy of imagery and the level of experienced vividness was reported on a scale from 1 to 7. If the reported accuracy was equal or lower than 5, the imagery task repeated until confidence level was improved. At the start and end of the session, participants took part in control conditions, in which the visual metronome was presented, but they were not asked to imagine the music. The control condition was performed in the two tempi of the melodies (73 and 107 BMP), while tapping to the beat ("control w/ tap") or keeping motionless ("control w/o tap").

When performed within a condition, the ISC method uses the participant's brain responses within a given brain area as a model to predict brain responses to the experience shared by all. When performed across conditions, the ISC method uses the averaged brain response to the experience in one condition as a model to predict brain responses to another experience in a different condition.

The ISC was performed within each of the following conditions, for each of the six melodies: Imagery w/o tap, control w/o tap, perception. ISC was also performed across the following conditions: Imagery w/o tap versus perception, imagery w/ tap versus perception, control w/o tap versus perception, control w/ tap versus perception. Within each of these comparisons, ISC was calculated between both matching and nonmatching melodies (or tempi) across the different conditions.

Inter-regional ISC Analysis

We calculated the inter-regional inter-subject correlation (IR-ISC, also known as inter-subject functional correlation) matrix between all parcels across brains of participants from two different conditions: perception versus imagery w/o tap, and perception versus imagery w/ tap. By measuring the inter-regional correlation across brains, instead of within a brain (common in functional connectivity analysis), the intrinsic neural correlations and nonneural confounds can be filtered out, thus

increasing the ability to detect inter-regional coupling induced by shared experiences across participants (Simony et al. 2016). In addition, by also comparing the responses across conditions, we can identify couplings driven by the commonalities across perception and imagery.

The neural signals X_i measured from participant i, $i = 1, \ldots, k$ are in form of a $p \times n$ matrix that contains signals from p neural sources (i.e., parcels) over n time points. All time courses were z-scored within participants to zero mean and unit variance. Thus, the participant-based IR-ISC was calculated by a Pearson correlation between single participant from one condition and the average of all participants from the other condition as follows: $\tilde{C}_i = \frac{1}{N} X_i [\frac{1}{k} \sum_i Y_j^T]$, where the neural signals Y_i measured from participant i, $i = 1, \ldots, k$. The cross-condition IR-ISC matrix was given by the following equation: $\tilde{C} = \frac{1}{k} \sum_i \tilde{C}_i$. The final IR-ISC matrix is given $(\tilde{C} + \tilde{C}^T)/2$, and then averaged with the final IR-ISC matrix calculated between individuals from the second condition and averaged signal from the first condition.

The IR-ISC was performed within each melody across the following conditions: Imagery w/o tap versus perception, imagery w/ tap versus perception, control w/o tap versus perception, and control w/ tap versus perception (Supplementary Fig. 1).

ISC Bootstrapping and Phase-randomization

The statistical likelihood of each observed correlation was assessed using a bootstrapping procedure based on phase randomization, to preserve the long-range temporal autocorrelation in the BOLD signal (Zarahn et al. 1997). The null hypothesis was that the BOLD signal in each area in each individual at any point in time (i.e., that there was no ISC between any pair of participants). For all conditions, phase randomization of each parcel time course was performed by applying a fast Fourier transform to the signal, randomizing the phase of each Fourier component, and inverting the Fourier transformation. This procedure scrambles the phase of the BOLD time course but leaves its power spectrum intact. For each randomly phase-scrambled surrogate dataset, we computed the ISC for all areas in the exact same manner as the empirical cross-condition correlation maps described above. That is, for ISC within condition, the Pearson correlation was calculated between that parcel's BOLD time course in one individual and the average of that parcel's BOLD time courses of all other individuals. For ISC between conditions, the Pearson correlation was calculated between that parcel's BOLD time course in one individual from one condition and the average of that parcel's BOLD time courses of all individuals from the other condition. The resulting correlation values were averaged within each parcel across all participants, creating a null distribution of average correlation values for all parcels.

To correct for multiple comparisons, we selected the highest ISC value from the null distribution of all parcels in a given iteration. We repeated this bootstrap procedure 10 000 times to obtain a null distribution of the maximum noise correlation values for each melody (i.e., the chance level of receiving high correlation values across all parcels in each iteration), and then averaged each iteration value across all melodies to obtain a null distribution for the averaged ISC maps. Familywise error rate (FWER) was defined as the top 5% of the null distribution of the maximum correlations values exceeding a given threshold (R*), which was used to threshold the veridical map (Nichols and Holmes 2001). In other words, in the ISC map, only parcels with mean correlation values (R) above the threshold derived from the bootstrapping procedure (R*) were considered significant after correction for multiple-comparisons and were marked as such on the final map.

Discriminability Analysis

To test for the discriminability of the temporal responses to each of the melodies, we used a permutation analysis (Kriegeskorte et al. 2008) that compares the ISC between matching melodies against the nonmatching melodies (Baldassano et al. 2018). In each parcel in the brain, significant discriminability was determined by resampling correlations of nonmatching melodies to generate a null distribution of the average across participants. That is, baseline correlations were calculated from nonmatching melody pairs, and compared to the average correlations of the matching melody pairs. This analysis ensures that the discovered temporal response patterns are content-specific at the melody level, as the correlation between neural patterns during matching melodies must on average exceed an equal-sized random draw of correlations between nonmatching melodies to be considered statistically significant. This procedure was performed for each parcel that was found significant in the ISC bootstrap analysis (see *ISC bootstrapping and phase-randomization*). The results were corrected for multiple comparisons across all significant parcels using False Discovery Rate (FDR; using *q*criterion of 0.05; Benjamini and Hochbert 1995).

This discriminability analysis was performed on the ISC values calculated between the perception and imagery conditions, as well as on the ISC calculated within the perception condition.

Classification Accuracy of Individual Melodies

We computed the level of discriminability of temporal patterns for individual melodies during perception, and across the perception and imagery conditions, in the parcels that showed significant ISC in the relevant condition. Participants were randomly assigned to one of two groups, an average time course was calculated within each group, and the data were extracted for each parcel of interest. When the classification was performed across conditions (e.g., perception and imagery), the data of each group were randomly sampled from one of the two conditions. Pairwise correlations were calculated between the two group means for all six melodies. For any given melody, the classification was labeled correct if the correlation with the matching melody in the other group was higher than the correlation with any other melody. Accuracy was then calculated as the proportion of melodies correctly identified out of six (chance level = 0.167). The entire procedure was repeated using 200 random combinations of the two groups sized $N = 12$ and $N = 13$. Statistical significance was assessed using permutation analysis in which for each combination of two groups, melody labels were randomized before computing the classification accuracy. Accuracy was then averaged across the 200 combinations and the procedure was performed 1000 times to generate null distributions for overall accuracy in a given parcel. Classification accuracy was corrected for multiple comparisons over all parcels of interest using FDR at threshold $q = 0.05$. The classification accuracy across the perception and imagery conditions was also calculated separately within each of the two tempo subgroups of the three melodies (chance level = 0.334).

IR-ISC Enhancement by Tapping

To measure the degree of neural response enhancement by tapping, we assessed the difference in reinstatement of the response to melodies during imagery when it was accompanied by tapping compared to motionless. Specifically, to measure effects of tapping on internal processing of melodies, in each brain area we subtracted the IR-ISC calculated between perception and motionless imagery, from the IR-ISC between perception and movement-accompanied imagery (i.e., IR-ISC(imagery w/ tap vs. perception) – IR-ISC(imagery w/o tap vs. perception)). Similarly, we also subtracted the IR-ISC calculated between perception and the control conditions (IR-ISC(control w/ tap vs. perception) – IR-ISC(control w/o tap vs. perception)). Significant enhancement of coupling across parcels was determined using a paired two-tailed permutation analysis that randomly assigned the correlations from the two imagery conditions (w/ or w/o tapping) to two new sham groups, generating a null distribution from the subtraction between the sham groups. We corrected for multiple comparisons across all parcels by controlling the FDR ($q = 0.05$).

Code Availability

Code supporting the findings of this study is available from the corresponding author upon request.

Results

Behavioral Assessment of Imagery Accuracy

Prior to the scans, participants completed two tests aimed at assessing the accuracy of their imagery of each of the learned melodies. The "Beat" test and the "Pitch" test were designed to capture temporal and pitch fidelity of the internally recalled melody, respectively (Herholz et al. 2008; Weir et al. 2015). Each test was performed under two movement instructions: during imagery, participants were either asked to be still or tap their finger to the rhythm of the metronome (see Methods, *Experimental design*).

In both the Beat and Pitch tests, average success rate across the six melodies was higher than chance level (Beat: M = 72.3%, SD = 11.6%; $t_{(24)} = 9.62$, $P \ll 0.0001$, $d = 3.93$; Pitch: M = 96.7%, SD = 2.4%; $t_{(23)} = 94.97$, $P \ll 0.0001$, $d = 39.6$), demonstrating the participants' ability to accurately internally recall the learned melodies (Supplementary Fig. 2).

Next, we explored whether the accuracy of imagery was affected by rhythmic tapping. Average success rate was similar whether participants tapped or not during imagery in both the Beat test (tap: M = 72%, SD = 11.1%, no tap: M = 72.7%, SD = 15%; $t_{(24)} = -0.27$, $P = 0.79$, $d = -0.11$) and the Pitch test (tap: M = 96.8%, SD = 2.1%, no tap: M = 96.6%, SD = 3.4%; $t_{(23)} = 0.36$, $P = 0.72$, $d = 0.15$). With the confirmation of participants' ability to accurately imagine the music, we examined whether there was neural reinstatement during imagery.

Similar Temporal Patterns between Participants during Imagery

We began by establishing to what extent, and where in the brain, imagining the melodies elicited synchronized neural response across people. We identified the brain areas that responded reliably across participants who imagined the same melody (without tapping) by calculating the temporal ISC (Hasson et al. 2004) within each of 400 cortical parcels (Schaefer et al. 2018) across brains (Fig. 2A). The response reliability map of imagery averaged across all six melodies is shown at $r > 0.14$ for visualization purposes, and an FWER was used to correct for multiple comparisons (Fig. 2B).

We found similar temporal responses across participants during imagery of each of the melodies in early and associative auditory cortices (Fig. 2B). These areas included parcels at the middle of the superior temporal plane (mSTP) bilaterally, which largely overlap with loci of early auditory processing in Heschl's gyri (Supplementary Table 1), as well as in more anterior and posterior auditory areas in the STP, and in lateral parts of the right superior temporal gyrus (STG). Most of these sensory regions also showed similar temporal response across participants during the perception of the melodies (Supplementary Fig. 3A, B).

To exclude the possibility that the observed similarities in the temporal response patterns were driven by the visual metronome (which was present in the imagery condition) or acoustic scanning noise, rather than the imagery task itself, we performed the same ISC analysis in the control condition (Fig. 2C), in which participants were exposed to the same sensory input as in the imagery condition (a bouncing ball in one of two tempi and sounds associated with the EPI sequence) but were not asked to imagine the melodies. Although some involuntary imagery might have taken place during the control condition, this analysis nonetheless revealed no significant correlations in the auditory cortices (Fig. 2D), thus confirming that the similar temporal patterns across people are driven by the mental imagery task, rather than the visual input.

Similar Temporal Pattern across Perception and Imagery

The preceding ISC in responses during imagery could be attributed to any of several different components of the imagery task. It might have been evoked by a common reinstatement of the previously perceived experience of music, but could also be due to nonspecific shared processes that are idiosyncratic to imagery, which are not manifested during music perception (e.g., certain control mechanisms or systematic modified mnemonic representation of the melodies). To investigate this question, we next looked for areas that specifically took part in the reinstatement of the perceptual experience of the melodies during imagery.

Although spatial overlap between areas that are activated during both auditory imagery and perception suggests some level of shared neural processing, a stronger form of shared neural representation is indicated when a region responds with the same temporal response profile to imagined and perceived forms of the same specific musical content. To test for direct correspondence between perception and imagery, we correlated the response time courses within each parcel when a given melody was perceived, to the response time courses when it was imagined. This analysis was performed on a between-participant basis by comparing each participant's response during imagery to that of all others during perception, which, when averaged, served as a group template of the typical response pattern for the perceptual experience (Honey et al. 2012; Chen et al. 2017; Nguyen et al. 2019; Regev et al. 2019) (Fig. 3A). This analysis revealed that in the bilateral early auditory areas (mSTP) and other higher auditory areas in the right STP and STG (Fig. 3B), temporal response patterns obtained during imagery of a given melody were similar to the patterns in other individuals during

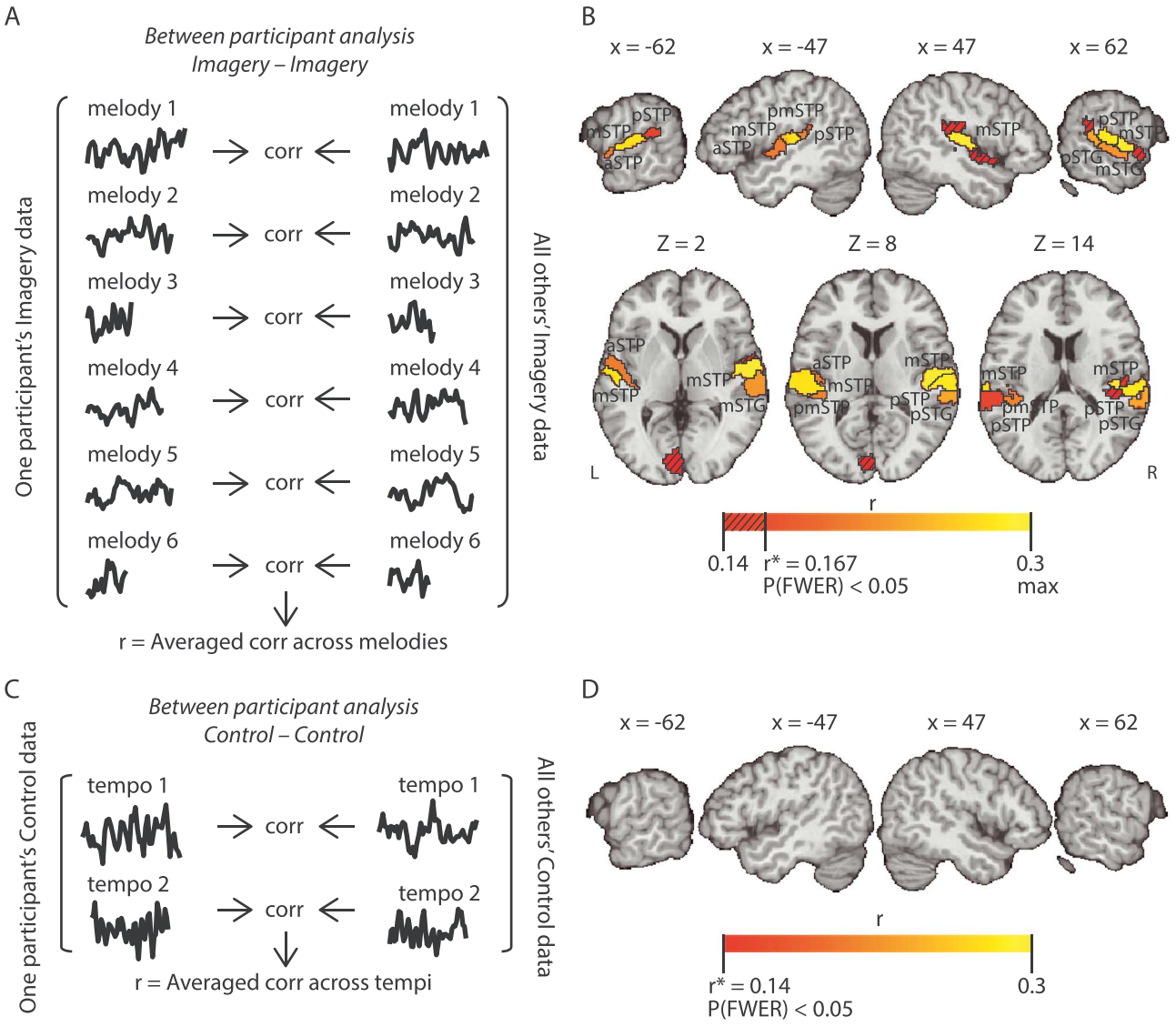

Figure 2. Temporal pattern similarity in the imagery and the control conditions (w/o tap). **A,** Schematic for between-participant analysis for the imagery condition. BOLD time courses were correlated between participants (each participant vs. the average of all others) for each of the six melodies in each parcel, to produce an averaged ISC during the imagery task. **B,** Cortical parcels where the highest temporal pattern similarity was observed across participants during imagery of the same melodies ($P_{(FWER)} < 0.05$; threshold $r = 0.14$ for visualization purposes). Regions in the lateral STG include middle and posterior parcels (mSTG, pSTG); Regions in the STP include middle, posterior, anterior, and posterior-medial parcels (mSTP, pSTP, aSTP, pmSTP). **C,** Schematic for between-participant analysis for the control condition, same as A, except that correlations were computed for each of the two types of task tempi. **D,** No cortical parcels in the temporal cortex show significantly similar temporal patterns between participants ($P_{(FWER)} < 0.05$).

perception of that same melody. Thus, the results indicate a shared neural response across perception and imagery, which is locked to the temporal dynamic of the melodies and consistent across individuals.

A comparison of the response patterns in perception to the responses in the control condition revealed no significant similarities in the auditory cortices (Supplementary Fig. 4), suggesting that the mere presence of the visual rhythm of the metronome did not account for the reinstatement of the neural patterns during imagery.

Does high similarity in temporal responses between imagery and perception of melodies reflect an accurate internal representation of the melody? To answer this question, we compared each individual's magnitude of inter-subject imagery-to-perception similarity, averaged across songs, to their general score in the behavioral Beat test of imagery accuracy. The correlation between these two scores revealed the strongest associations to be in the right middle and posterior STG ($r = 0.4$, $P = 0.047$ and $r = 0.43$, $P = 0.037$, uncorrected for multiple comparisons; Supplementary Fig. 5), out of the all parcels where the similarity of response between imagery and perception was significant. This might indicate that the reinstatement of neural responses during imagery in these regions in the right STG can predict the accuracy of the internal musical experience.

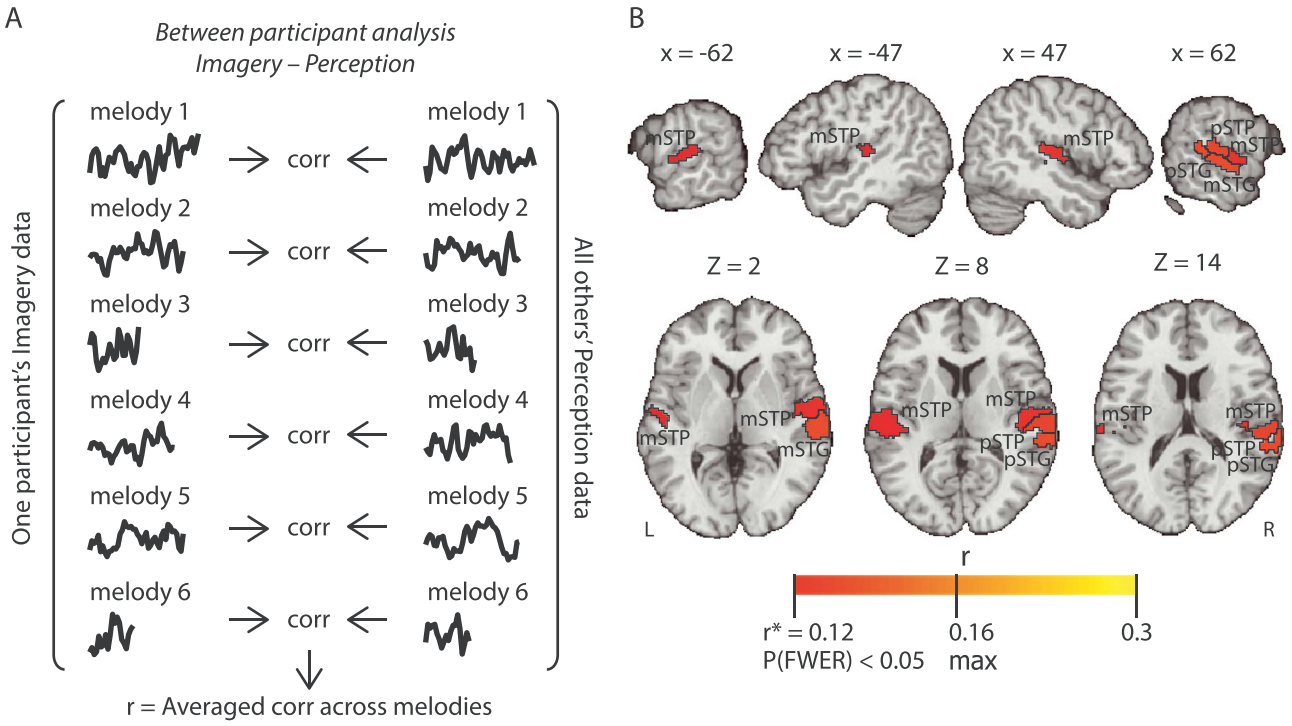

Figure 3. Temporal pattern similarity between imagery (w/o tap) and perception. **A,** Schematic for between-participant similarity analysis across imagery and perception. The correlations were computed between every matching pair of imagery/perception melodies across participants. **B,** Cortical parcels where the highest temporal pattern similarity was observed across participants between imagery and perception of melodies ($P_{(FWER)} < 0.05$).

Neural Reinstatement of Melody-Specific Temporal Patterns

The consistency in temporal response pattern across perception and imagery could potentially represent general cognitive trends common in both tasks, regardless of the identity of the melody (e.g., rapid buildup of attention). At the same time, the similarities between perception and imagery could also represent the reinstatement of the specific musical content, and be driven by temporal response patterns that are unique to the experience of each melody. To test whether the reinstated responses are discriminable from each other, we used a permutation analysis (Kriegeskorte et al. 2008; Baldassano et al. 2018) that compares neural pattern similarity between matching melodies against the nonmatching melodies, across perception and imagery (see Methods, *Discriminability analysis*). This analysis reveals regions containing reinstated melody-specific patterns, as statistical significance is only reached if matching melodies (same melody in perception and imagery) can be differentiated from nonmatching melodies (Fig. 4A, gray baseline). The discriminability analysis was performed in each of the five parcels in the STP and STG that showed a significantly similar response pattern across perception and imagery. The results of this analysis showed a within-vs-between melody difference in the right mSTG ($P < 0.005$, FDR-corrected $q = 0.05$), and a smaller difference in the right pSTG (did not survive correction; Fig. 4A). In other words, the similarity in temporal response patterns between perception and imagery in the right STG represents reinstatement of the specific musical content. These specific patterns are not unique to the internal experience of individuals but are shared across all participants who imagine the same musical content.

To further explore *how* discriminable the reinstated neural patterns of melodies were from each other, we performed a classification analysis (Chen et al. 2017) for the parcels that showed melody-specific response patterns (i.e., right mSTG and pSTG). Participants were randomly assigned to one of two groups ($N = 12$ and $N = 13$), and an average time course for each melody was calculated from the perception condition in one group and from the imagery condition in the other group. Classification accuracy across groups was calculated as the proportion of melodies correctly identified out of six. The procedure was repeated using 200 random combinations of groups and averaged. Statistical significance was assessed using a permutation analysis in which the melody labels were randomized before calculating the accuracy. Compared with the previous discriminability analysis, while the use of averaged group time courses may increase the power of the classification accuracy, this analysis is also stricter in its binary definition of a correct identification of a melody (see Methods, *Classification accuracy*).

The classification accuracy was found to be significant in both the right mSTG and pSTG, with 35% and 26% correctly labeled melodies respectively (Fig. 4B; chance level 16.7%, $P < 0.001$; FDR-corrected $q = 0.05$). Since the two different tempi of the melodies may have contributed to the discriminability of the neural patterns between melodies, we repeated the classification analysis within each tempo subgroup of three melodies. The classification accuracy was again found to be significant within each of the two tempo subgroups at the right mSTG

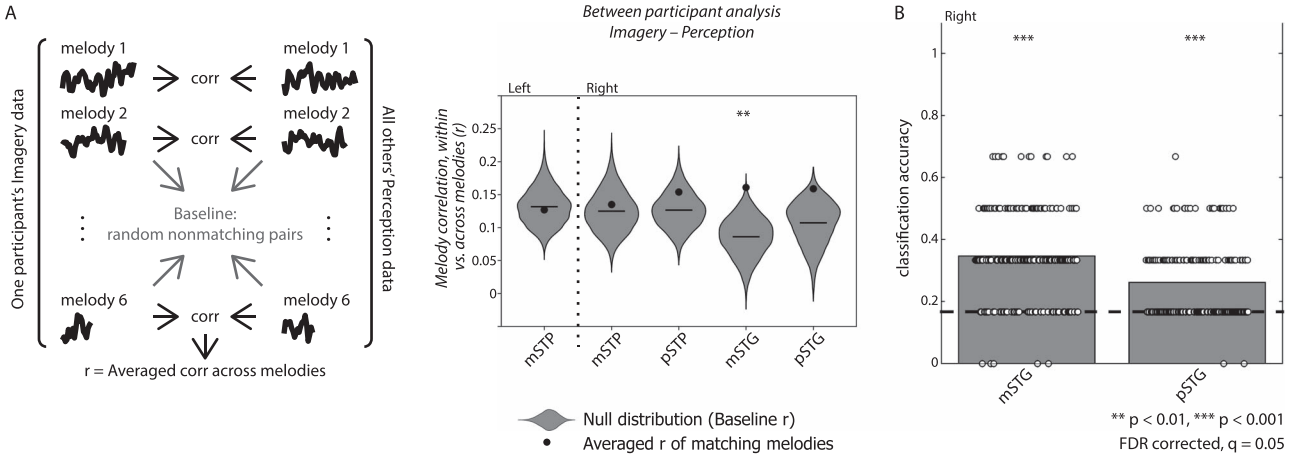

Figure 4. Neural reinstatement of melody-specific temporal patterns. **A,** Discriminability of temporal patterns in parcels that showed similarity across imagery and perception. The correlations were computed between every matching (black) and nonmatching (gray) pair of imagery/perception melodies across participants. Reinstated temporal patterns were discriminable from each other when the similarity between matching melodies was greater than between nonmatching melodies. The black circles show average correlation of matching melodies. Statistical significance was determined by generating a null distribution of random pairs of nonmatching melodies (gray violin baseline; FDR-corrected q = 0.05). **B**, Classification accuracy of melodies across brains between imagery and perception in the right mSTG and pSTG. Participants were randomly assigned to one of two groups (N = 12and N = 13), and an average time course calculated from the imagery condition in one group and from the perception condition in the other for all six melodies. Accuracy was calculated as the proportion of melodies correctly identified out of six. The entire procedure was repeated using 200 random combinations of the two group sizes (white circles), and an overall average was calculated for each parcel (right mSTG 35%, pSTG 26%, gray bars; chance level 16.7%, dashed line; FDR-corrected q = 0.05).

(BPM 73: 44%, BPM 107: 45%, $P < 0.001$; chance level: 33%), but only in one subgroup at the right pSTG (BPM 73: 41%, $P < 0.001$; BPM 107: 34%, $P = 0.32$; Supplementary Fig. 6A, B). Overall, this suggests that the right mSTG, and perhaps less so the right pSTG, reinstate melody-specific temporal neural patterns during imagery.

The Effect of Rhythmic Tapping on Reinstatement of Melody-Specific Temporal Response Patterns

Do neural representations of melodies change when imagery is accompanied by rhythmic movement? To address this question, we examined the effect of rhythmic tapping to the beat of imagined melody on the reinstatement of its unique response pattern, and on the propagation of the imagined content across functional networks.

For this purpose we began by repeating the ISC analysis, which compared the response time courses when a melody was perceived, to the response when it was imagined, but with accompanying rhythmic tapping movements (Fig. 5A). The analysis revealed similar temporal response patterns between perception and imagery, which extended to further auditory areas, beyond parcels identified in the previous comparison to motionless imagery. Specifically, more parcels in the STP bilaterally and in the left lateral STG showed significant response similarities across perception and imagery of melodies due to the added motor component to the imagery task (Fig. 5B).

The introduced synchronized movements also enhanced the reinstatement of melody-specific response patterns in the auditory cortex. The discriminability analysis was performed using the movement-accompanied imagery data in the nine parcels in the STP and STG that showed significantly similar responses across perception and imagery. We identified significant melody discriminability in regions that also exhibited such response patterns during motionless imagery (right mSTG $P < 0.005$, right pSTG $P < 0.05$; Fig. 5C), as well as an increase in classification accuracy (right mSTG: 46%, $P < 0.001$; $t_{(398)} = 7.77$, $P \ll 0.0001$, $d = 0.78$; right pSTG: 40%, $P < 0.001$; $t_{(398)} = 10.33$, $P \ll 0.0001$, $d = 1.04$; Fig. 5D). Furthermore, the discriminability extended to further regions in the STP bilaterally, including the aSTP parcels (left $P < 0.005$, right $P < 0.0001$) and left pmSTP ($P < 0.001$), which also showed significant classification accuracy (left aSTP 45%, right aSTP 70%, and left pmSTP 36%, $P < 0.001$). All these regions, except the right pSTG, showed significant classification accuracy as well, when calculated within each of the different tempo subgroups (Supplementary Fig. 6C, D). This suggests that temporal patterns of brain activity in the associative auditory cortex (especially parts of the STP) were modified by the incorporation of synchronized tapping with imagery in a consistent manner across individuals, in becoming more similar to responses during perception, and more distinguishable across melodies. Thus, tapping did not just improve the general beat representation of the melodies but also strengthened the unique neural representation of melodies with the same tempo.

The Effect of Rhythmic Tapping on Propagation of Musical Content across Cortical Areas during Imagery

What functional connections in the brain might support the observed spread of melody-specific information during rhythmic movement? To map how melody-locked neural responses are shared *across* pairs of brain areas, we used an inter-subject modification to the common functional connectivity analysis (Simony et al. 2016). In this analysis, the correlation of temporal responses was performed not only on a between-participant basis and across the perception and imagery conditions, but also across cortical parcels. For example, we can test whether the response to an imagined melody in the auditory cortex of one participant is correlated with the response to that melody in the precentral gyrus when that melody is perceived by another participant. By comparing the responses across conditions, we

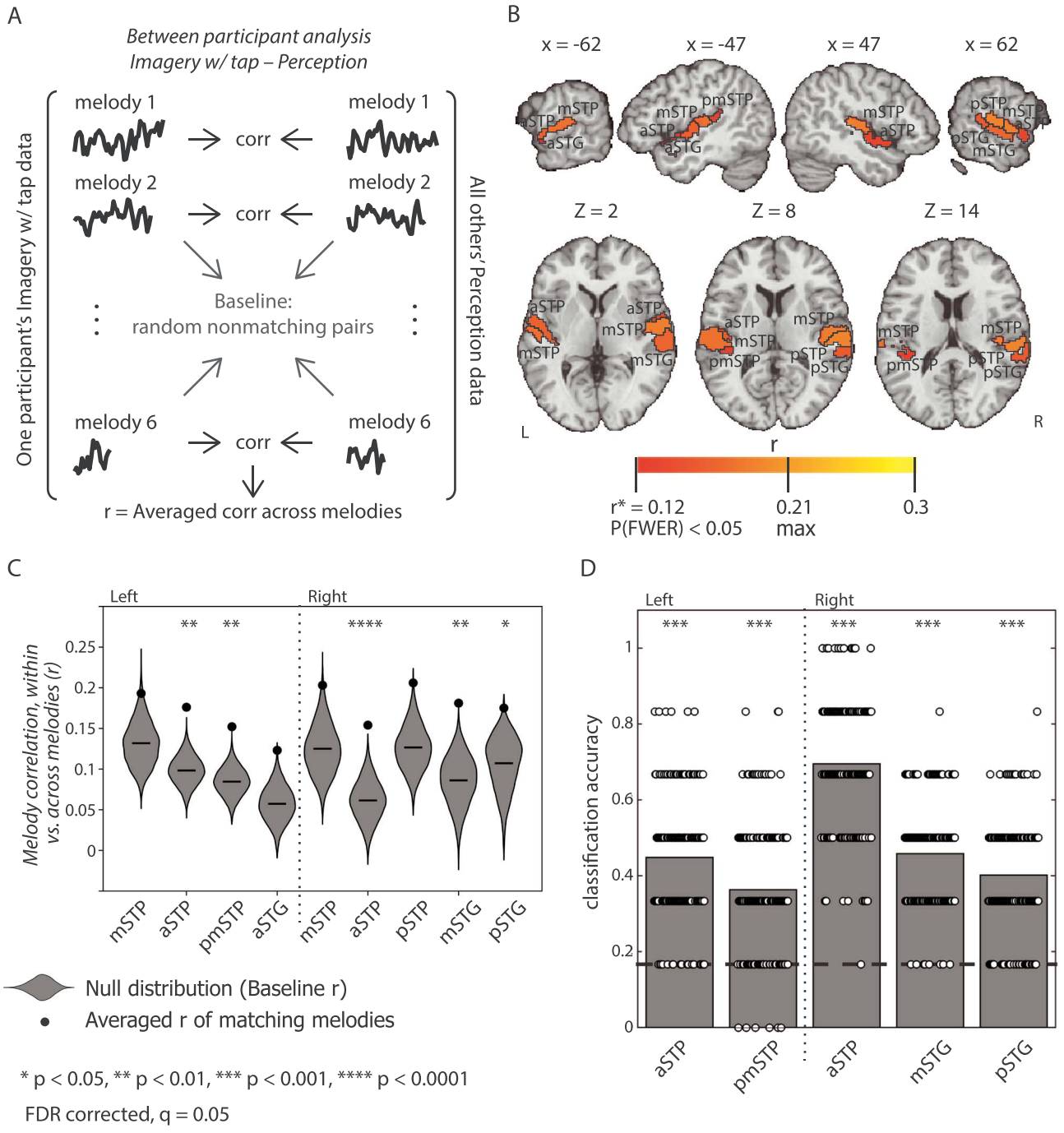

Figure 5. Temporal pattern similarity between imagery accompanied by tapping and perception. **A,** Schematic for between-participant similarity analysis across imagery w/ tap and perception. Same as Fig. 3A, except that correlations were computed using the imagery condition that included tapping. **B,** Cortical parcels where the highest temporal pattern similarity was observed across participants between imagery w/ tap and perception of matching melodies ($P_{(FWER)} < 0.05$). **C,** Discriminability of temporal patterns in parcels that showed similarity across imagery w/ tap and perception (FDR-corrected q = 0.05) (see Fig. 4A). **D,** Classification accuracy of melodies across brains between imagery w/ tap and perception (right mSTG 46%, pSTG 40%, aSTP 70%, left aSTP 45%, pmSTP 36%, gray bars; chance level 16.7%, dashed line; FDR-corrected q = 0.05) (see Fig. 4B).

can identify inter-regional couplings driven by the commonalities across perception and imagery, such as the internal representation of musical experience. This analysis was performed for each melody between the perception and each of the two imagery conditions (w/ or w/o tapping) across all 400 parcels, which were organized in resting-state functional networks (Yao et al. 2011; Schaefer et al. 2018; Supplementary Fig. 1A, B).

To describe the degree of inter-regional coupling enhancement by tapping, we calculated the difference of the mean correlation during motion-accompanied imagery from the mean

correlation during motionless imagery. The more positive the resulting value, the more reliable the inter-regional responses while tapping, whereas negative values indicate that the responses are more reliable without tapping, and a value close to zero indicates that the neural responses to the music during imagery are similar with and without rhythmic movement. Significant enhancement of coupling was determined using a (paired two-tailed) permutation analysis that randomly assigned the correlations from the two imagery conditions (with or without tapping) to two new sham groups, generating a null distribution from the subtraction between the sham groups. We corrected for multiple comparisons across all 400 parcels by controlling the FDR ($q = 0.05$).

Melody-locked inter-regional responses were most strongly enhanced between parcels of the ventral somatomotor (VSM) network when tapping accompanied imagery (M = 0.048, SD = 0.017; Fig. 6A, pink). In addition to its somatomotor regions, this functional network also encompasses the bilateral STP (Fig. 6B), including the parcels in which melody-specific reinstatement was extended to when imagery was accompanied by tapping (i.e., right aSTP and pmSTP, left aSTP; see Fig. 5). However, a closer look at this network revealed that the response enhancement was not limited to these auditory parcels but also extended between the STP and other parcels of the network, including insular and ventral somatomotor areas (Fig. 6C).

Furthermore, tapping enhanced melody-locked inter-regional responses between the VSM network – especially the STP – and other functional networks. We observed the strongest increase in inter-network coupling with the temporo-parietal network (M = 0.036, SD = 0.016; Fig. 6A, C), including parcels in the right STG that showed melody-specific reinstatement during imagery (Fig. 4). The next strongest increase in inter-network coupling was observed with the dorsal somatomotor (DSM) network (M = 0.033, SD = 0.021; Fig. 6A, C), which was previously found to accommodate the hand and foot regions (Yao et al. 2011). Other strong coupling enhancements were observed between the VSM network, especially the STP, and cognitive networks, such as one of the ventral attention networks (M = 0.03, SD = 0.019), as well as with the peripheral visual cortices (M = 0.025, SD = 0.015) (Fig. 6A).

To further test whether the more reliable inter-regional responses while tapping during imagery were indeed driven by the musical experience common across perception and imagery, we repeated the inter-regional analysis using the data from the control conditions instead of the imagery conditions (Supplementary Fig. 1C, D). The subtraction between the correlations during motion-accompanied and motionless control showed a relatively weak enhancement in inter-regional coupling during tapping across all functional networks (e.g., in VSM M = 0.004; Fig. 6D). These results demonstrate that the enhancement in coupling during tapping-accompanied imagery is not likely to be an artifact simply due to the inclusion of motion in the task, but rather represents an increase in the propagation of information related to the musical imagery task itself.

Discussion

We found that the specific contents of consciousness during auditory imagery can align the neural responses across individuals in both early and associative auditory cortices (Fig. 2). In a subset of these regions, we found that the temporal response patterns recorded during the perception of music were reactivated during imagery in a melody-specific manner, indicating that the response is specific to the imagined content. This reinstated and shared brain activity was observed in the middle of the right STG, in the absence of any auditory sensory cues (Figs 3 and 4). Furthermore, when music imagery was accompanied by synchronized rhythmic tapping, the distinctive responses to the melodies were reinstated in additional associative auditory areas of the superior temporal plane (Fig. 5), thus increasing the resemblance to the neural representation of the perceived stimulus. This extended reinstatement was also accompanied by an increase in the spread of the melody-locked response between the STP and varied functional networks, most prominently with somatomotor areas (Fig. 6), all of which supports an important role for motor-to-auditory influences in enhancing the fidelity of internal auditory representations.

Neural Representation of Specific Imagined Musical Content

Prior studies reported spatial overlap between areas that respond to perceived and imagined music (e.g., Halpern and Zatorre 1999; Hickok et al. 2003; Kraemer et al. 2005; Meyer et al. 2007; Herholz et al. 2012). However, such overlap does not reveal whether the region contains information about the specific ongoing auditory content being retrieved and experienced. The present study goes beyond previous findings by demonstrating that brain activity time courses when listening to real-life complex music were reinstated during imagery of the music in the STP and STG (Fig. 3B), and that these reinstated melody patterns were discriminable from one another in the right STG (Fig. 4A, B). Moreover, our methodology revealed the extent to which the subjective internal experience of hearing music in the "mind's ear" has shared neural representations across people, even though each person learned the melodies as they chose, and had different levels of musical expertise.

The similarity in temporal responses across perception and imagery was observed in early auditory areas bilaterally (Fig. 3B). Previous studies have found a spatial overlap in neural activity in secondary and associative areas, but evidence for shared activity is scarcer for primary areas (e.g. Griffiths 2000; Halpern et al. 2004; Bunzeck et al. 2005; Herholz et al. 2012, but see Yoo et al. 2001; Kraemer et al. 2005; Oh et al. 2013). While between-participant synchronization in the early auditory cortex has been reported during perception (Lerner et al. 2011; Honey et al. 2012; Regev et al. 2013; Farbood et al. 2015), this is the first study to capture temporal alignment in this sensory area during imagery, a purely internal process without any externally driven acoustic component. Although we observed similarities in parcels that overlap with Heschl's gyrus, the whole-brain parcellation used here (based on resting-state functional connectivity; Schaefer et al. 2018) does not allow an accurate dissociation between primary and secondary regions. Therefore, it is possible that the recorded similarities in the early auditory cortices are driven by activation in the secondary, rather than primary auditory regions.

Despite the similarity between perception and imagery in the early auditory areas, we did not identify a reinstatement of melody-specific responses during imagery in these areas, but rather in lateral areas of the STG (Fig. 4). This result is consistent with previous neuroimaging studies that attempted to decode environmental sounds across perception and imagery using multivoxel pattern analysis, but did not find distinctive representation in early auditory areas (Meyer et al. 2010; Vetter et al. 2014; de Borst et al. 2016; Gu et al. 2019). Our results suggest that

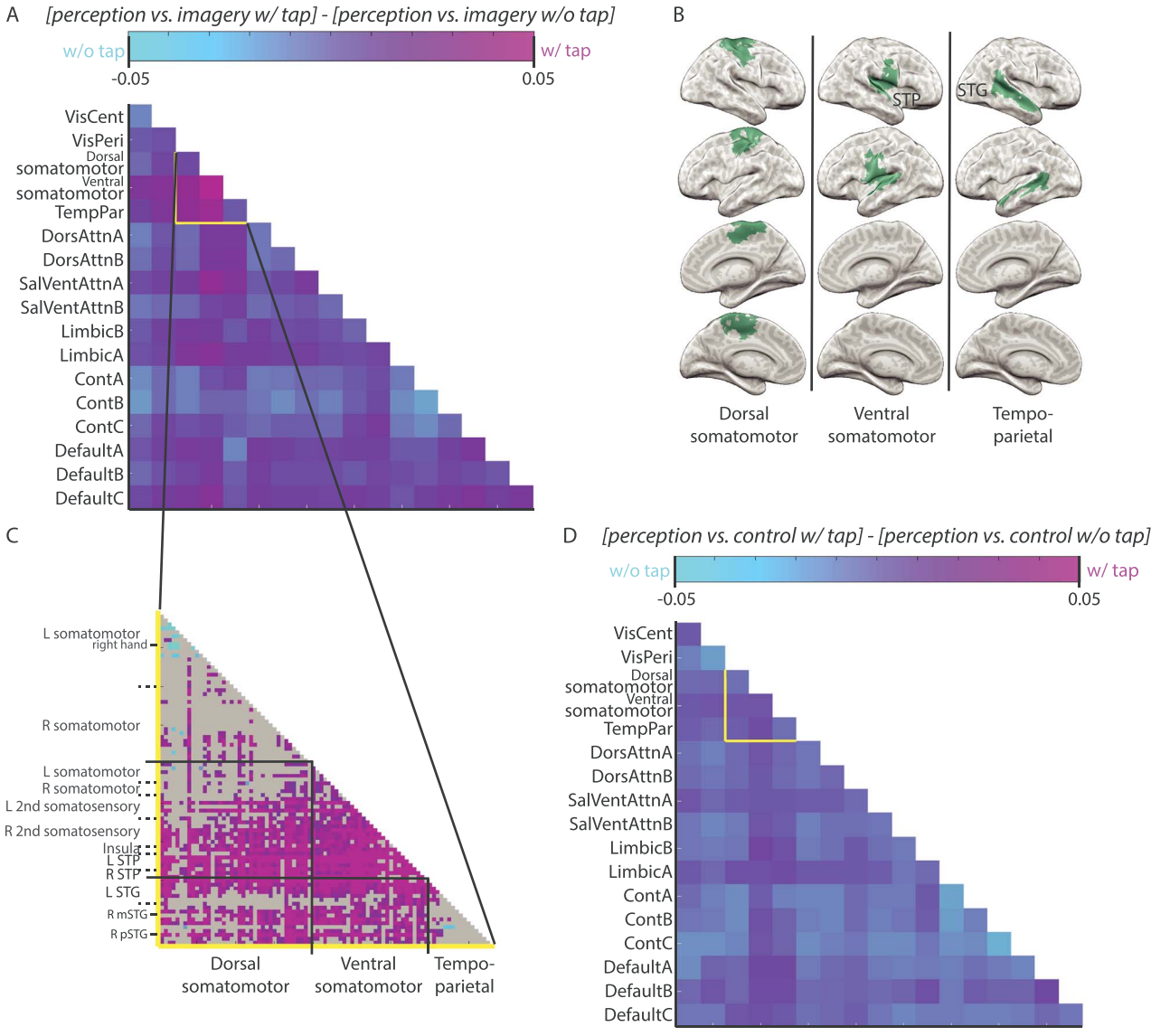

Figure 6. Rhythmic tapping during music imagery enhanced inter-regional melody-locked responses across functional networks. Degree of enhancement was assessed as the difference in level of melodies' reinstatement when imagery was accompanied with tapping and was motionless. **A,** Inter-regional coupling in reinstatement of music was enhanced across several functional networks. **B,** Anatomical localization of three independently defined functional networks that showed the strongest average inter-regional enhancement: the VSM, the tempo-parietal, and the DSM networks (Yao et al. 2011). **C,** The strongest enhancements were observed between parcels in the VSM network that includes the STP, as well as ventral somatomotor areas. Other strong enhancements were observed between this network and parcels in the tempo-parietal and DSM networks, including the lateral STG and parcels which corresponds to hands topography (FDR-corrected $q = 0.05$). **D,** The enhancement of inter-regional coupling by tapping was low when assessed as the difference between the control conditions w/ and w/o tap.

the temporal response similarity observed in the early auditory region across perception and imagery might represent general operations shared by both tasks, with temporal dynamics that are not specific for each melody (but might still influence the auditory experience).

Previous studies have shown that during processing of real-life auditory stimuli, such as music and speech, activity time courses in associative auditory regions are synchronized across individuals and locked to mid-level structures in the stimulus (e.g. sentences and musical phrases; Lerner et al. 2011; Farbood et al. 2015). The current study extends these findings by showing that synchronized neural responses during perception can give rise to shared neural responses during imagery in the right lateral associative auditory cortices, reinstated at will from memory (Fig. 4). The shared neural response across people was robust enough to overcome any potential imprecisions that might have incidentally occurred during the imagery task.

Although such neural reinstatement during auditory imagery was never captured before, this finding corresponds well with previous studies that revealed neural organization in high-level visual areas for recalled memories of specific movie scenes, which is similar to perception and sometimes common across participants (Buchsbaum et al. 2012; Chen et al. 2017; Zadbood et al. 2017). In agreement with our findings in the auditory

modality, the key role of the right auditory cortex in imagery of music has been described before in temporal-lobe patients and noninvasive neuroimaging studies (Zatorre and Halpern 1993; Halpern and Zatorre 1999; Meyer et al. 2007; Ding et al. 2019), and is also in keeping with an extensive body of evidence supporting the idea that melodic information is better represented in auditory areas within the right vs. left hemisphere (Zatorre et al. 2002; Albouy et al. 2020).

In general, the responses observed here during either imagery or perception were limited to the temporal lobe. During imagery of music, neural activity has previously been described also in frontal regions and most frequently in medial and lateral premotor cortices (Hickok et al. 2003; Leaver et al. 2009; Herholz et al. 2012; Jacobsen et al. 2015; Li et al. 2020). In the majority of these studies the imagined music had associated lyrics that might have had an important role in the activation of motor regions, as has been observed also during inner vocalization of speech (Tian et al. 2016). We did not observe any significant response in these areas, perhaps because participants were asked to plainly imagine the instrumental music rather than internally vocalize it. The lack of reliable dynamic responses across people does not necessarily imply that there was no neural activity in these regions, however, only that this pattern of activity was not time-locked to the common imagined content. It is possible that additional regions were active during imagery and perception, but did not respond dynamically to the presented or retrieved information (for example, they could have maintained a constant level of activity over time, related to more general cognitive or sensory processes).

Unlike the current study, reliable inter-subject temporal response during music perception was previously captured in structures in the frontal and parietal lobes (Alluri et al. 2012; Abrams et al. 2013; Farbood et al. 2015). However, these structures varied across experiments, and the response reliability there was weaker than described for the temporal cortex. This discrepancy with our results, which showed mainly the superior temporal cortex (Supplementary Fig. 3B), could be a result of differences in stimulus and task demands. For instance, we used shorter musical segments (a single minute instead of longer melodies), which was essential for allowing their memorization, but cannot reveal areas that encode information integrated over longer periods of time (Farbood et al. 2015).

The Effect of Rhythmic Movement on the Representation of Imagined Music

Temporal response patterns in the auditory cortex during musical imagery became more similar to the responses during perception when rhythmic tapping accompanied the imagery. Since the only shared experience across imagery and perception was the representation of music, and not the visual metronome or rhythmic movement (which were present only during imagery), the observed response similarities cannot directly represent any potential propagation of the rhythmic oscillations from visual or motor cortices into the auditory cortex. In addition, when the responses during perception were compared with a parallel control condition to imagery that included tapping to the metronome without an internal representation of music, no significant signal correlations were observed in the brain (Supplementary Fig. 4). Thus, although the spread of shared responses further into the STP could have been facilitated by a rhythmic modulation of the auditory cortex (e.g. through entrainment), this simple rhythmic component is not the source of the observed cross-condition similarities.

The strongest increase in neural similarity to perception and melody-specific reinstatement was observed in bilateral associative auditory areas in the STP, anterior and posterior to early auditory cortex (Fig. 5). These associative auditory areas in the temporal plane might differ from those in the lateral STG in their involvement in integration and recognition of relatively more fine-grained auditory patterns (Binder et al. 2000; Woods et al. 2009; Norman-Haignere et al. 2015; Hamilton et al. 2020), yet the distinct functional role of these regions in humans is still under debate. Thus, the movement might have enhanced various aspects of sensory processing that take part in both perception and imagery of music, such as spectrotemporal tuning of the neural receptive fields (Martin et al. 2018), or the sharpening of temporal selection of relevant auditory information (Nozaradan et al. 2016; Morillon and Baillet 2017).

The increase in the similarity of the responses between perception and imagery and in the reinstatement of specific melodies could have been driven by modulation of different components of the temporal signal. For instance, the tapping manipulation might have had a broad effect on the memory of the melody as a whole, making the neural response during imagery more similar to perception in a uniform manner over time. Alternatively, the enhancement might have had a more oscillatory profile, with focal increases in similarity to perception around the co-occurring tapping events (Morillon and Baillet 2017). Furthermore, the increase in similarity of the responses during imagery to those during the original sensory experience might have captured different enhancements in the internal representation of music. It could have been a result of more accurate neural representation of certain auditory features, instead of or in addition to participants' possible improved ability to align their imagery to the metronome. If only the latter is true, ISCs might have increased without any rise in a population's SNR of the perceptual representation. In this hypothetical case, our results would suggest that the added rhythmic movement had an ability to increase the temporal accuracy of retrieval of music, compared to a unimodal visual presentation of rhythm (as introduced by the visual metronome).

The strong anatomical and functional connection between the somatomotor and auditory cortices has been widely reported in variety of contexts, including both comprehension and production of complex auditory content (Hickok et al. 2003; Zatorre et al. 2007; Saur et al. 2008; Peretz et al. 2009; Rauschecker and Scott 2009; Hickok et al. 2011; Rauschecker 2011; Andoh et al. 2015). The strong interaction between these systems was specifically demonstrated in top-down somatomotor influences on the processing of perceived music, whether the movement generated the sounds (Zatorre et al. 2007; Repp and Su 2013) or just accompanied them (Nozaradan et al. 2016; Morillon and Baillet 2017). These influences were also demonstrated in the enhancing effect of articulatory movement on processing of the silently uttered speech (Okada et al. 2018; Zhang et al. 2020). To our knowledge, our results are the first to suggest that the motor system might have a broader modulatory role, not only during covert speech or when making sense of perceived sounds, but also when nonlinguistic auditory information is represented entirely internally.

We found that the extended neural reinstatement of melodies in the auditory cortex due to tapping was accompanied by increased interactions between somatomotor and auditory cortices. Tapping during imagery induced the strongest increase

in shared musical content between auditory parcels in the STP and somatomotor parcels around ventral parts of the central sulci (Fig. 6C). These parcels create together a functional network typical for resting state (Yao et al. 2011), thereby providing a potential avenue for auditory–motor interactions (such as the current motor modulation) during various tasks. Although the influence of tapping was most prominent within this network, it also encompassed interactions between auditory and motor regions in other functional networks, including parcels in the lateral STG and the dorsal somatomotor cortex. Furthermore, tapping during imagery also led to the spread of content-specific coupling between the STP and other neural systems, such as parts of the ventral attention network and visual cortices. This suggest that the effect of rhythmic movement on internal representation of music might be a result of a complex interplay between motor, sensory, and high-order networks.

Whereas it is commonly hypothesized that motor commands lead to better prediction of sensory inputs and thus benefit individual's adaptability to its environment (Wolpert et al. 1995; Schubotz 2007; Schroeder et al. 2010; Cannon and Patel 2020), the advantage of such a mechanism is not as clear during imagery, in which the sensory information had ostensibly already been established and can be retrieved at will. Nevertheless, the undeniable gap between stored knowledge and its transient representation in conscious experience could perhaps be somewhat mitigated by this mechanism. The modulatory capability of the motor system may be "repurposed" using deliberate motor strategies, in order to enhance the experience of covert auditory information, as has been shown before for overt speech and music (Kotz et al. 2009; Morillon et al. 2014; Schön and Tillmann 2015; Morillon and Baillet 2017). While in the current design tapping to the beat of the music was an imposed experimental manipulation, rhythmic movements often appear spontaneously when people imagine music, and silent articulation is common when seeking to establish internal speech during silent reading (McGuigan et al. 1964; Hardyck and Petrinovich 1970). Thus, movement may have an important role in promoting the conscious reinstatement of complex auditory content.

Although this melody-specific reactivation was enhanced by rhythmic movement, we could not identify a corresponding behavioral improvement in imagery accuracy. Our measurement of imagery accuracy was modeled after an established measurement of internal pitch and beat representation (Herholz et al. 2008; Weir et al. 2015), but it was heavily modified for the use of longer and more complex musical segments, which could have influenced our ability to capture differences in imagery accuracy across conditions. As a result, we are unable to say how the tapping may have enhanced the internal experience of music by the participants, and the exact manner in which the modification of neural activity corresponds with that experience.

Overall, by following the neural responses to the specific content of imagery, this study was able to shed light on a largely impenetrable mental construct. The results reveal the existence of common neural activation for complex and continuous auditory content in regions where encoded information is largely sensory, bridging perceptual and imagined experiences. Over and above individual differences, we were able to trace the neural responses associated with familiar musical segments, and show the common ability to reactivate them at will when bringing to mind musical content. The fact that these observations were made as individuals were engaged with complex musical content testifies to the robustness and ecological validity of the phenomena. We also demonstrate that the similarity in the neural codes elicited by perception and imagery is not limited to isolated neural populations, but could also be observed in comparable intricate interactions between neural modalities, such as motor influence on the auditory system. Future work can explore what aspects of musical sequences (e.g., timbral, textural, or rhythmic) are most correlated with melody-specific fMRI time series during imagery (Alluri et al. 2012), and can study whether the objective and subjective mental reinstatement of sounds may be modified by the spread of neural reinstatement due to rhythmic movement.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Notes

We thank Janice Chen and Benjamin Morillon for providing comments and suggestions, Vanessa Wong and the McConnell Brain Imaging Centre of the MNI for assistance in data collection, and to the Zatorre lab members for their support. *Conflict of Interest*: None declared.

Funding

This study was supported by a Canadian Institute for Advanced Research (CIFAR) Stimulus grant (to R.Z., A.D.P., and A.M.O.), a Foundation Grant from the Canadian Institute of Health Research (to RZ), and a postdoctoral fellowship from NSERC-Create in Complex Dynamics (to MR). The authors acknowledge the financial support of Health Canada, through the Canada Brain Research Fund, an innovative partnership between the Government of Canada and Brain Canada, and the Montreal Neurological Institute and from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative.

References

Abrams DA, Ryali S, Chen T, Chordia P, Khouzam A, Levitin DJ, Menon V. 2013. Inter-subject synchronization of brain responses during natural music listening. *Eur J Neurosci*. 37:1458–1469.

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP. 2013. Shared representations for working memory and mental imagery in early visual cortex. *Curr Biol*. 23:1427–1431.

Albouy P, Benjamin L, Morillon B, Zatorre RJ. 2020. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*. 367:1043–1047.

Alluri V, Toiviainen P, Jaaskeläinen IP, Glerean E, Sams M, Brattico E. 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*. 59:3677–3689.

Amit E, Greene JD. 2012. You see, the ends Don't justify the means: visual imagery and moral judgment. *Psychol Sci*. 23:861–868.

Andoh J, Matsushita R, Zatorre RJ. 2015. Asymmetric interhemispheric transfer in the auditory network: Evidence from TMS, resting-state fMRI, and diffusion imaging. *J Neurosci*. 35:14602–14611.

Baddeley A, Logie R. 1992. Auditory imagery and working memory. In: Reisberg D, editor. *Auditory imagery*. Hillsdale, NJ: Lawrence Erlbaum, pp. 179–197.

Baldassano C, Hasson U, Norman KA. 2018. Representation of real-world event schemas during narrative perception. *J Neurosci*. 38:9689–9699.

Ben-Yakov A, Honey CJ, Lerner Y, Hasson U. 2012. Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage*. 63:501–506.

Benjamini Y, Hochbert Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 57:289–300.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*. 10:512–528.

Böck S, Korzeniowski F, Schlüter J, Krebs F, Widmer G, eds. 2016. Madmom: a new Python audio and music signal processing library, in Proceedings of the 2016 ACM Multimedia Conference, Amsterdam, The Netherlands. 2016.1174–1178 p.

Brainard DH. 1997. The psychophysics toolbox. *Spat Vis*. 10:433–436.

Buchsbaum BR, Lemire-Rodger S, Fang C, Abdi H. 2012. The neural basis of vivid memory is patterned on perception. *J Cogn Neurosci*. 24:1867–1883.

Bunzeck N, Wuestenberg T, Lutz K, Heinze H-J, Jancke L. 2005. Scanning silence: mental imagery of complex sounds. *NeuroImage*. 26:1119–1127.

Cannon JJ, Patel AD. 2020. How beat perception co-opts motor neurophysiology. *Trends Cogn Sci*. 25:137–150.

Cervantes Constantino F, Simon JZ. 2018. Restoration and efficiency of the neural processing of continuous speech are promoted by prior knowledge. *Front Syst Neurosci*. 12:1–11.

Chang C, Cunningham JP, Glover GH. 2009. Influence of heart rate on the BOLD signal: the cardiac response function. *NeuroImage*. 44:857–869.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U. 2017. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci*. 20:115–125.

de Borst AW, Valente G, Jaaskeläinen IP, Tikka P. 2016. Brain-based decoding of mentally imagined film clips and sounds reveals experience-based information patterns in film professionals. *NeuroImage*. 129:428–438.

Ding Y, Zhang Y, Zhou W, Ling Z, Huang J, Hong B, Wang X. 2019. Neural correlates of music listening and recall in the human brain. *J Neurosci*. 39:8112–8123.

Dinstein I, Hasson U, Rubin N, Heeger DJ. 2007. Brain areas selective for both observed and executed movements. *J Neurophysiol*. 98:1415–1427.

Farbood MM, Heeger DJ, Marcus G, Hasson U, Lerner Y. 2015. The neural processing of hierarchical structure in music and speech at different timescales. *Front Neurosci*. 9:157.

Foster NEV, Halpern AR, Zatorre RJ. 2013. Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*. 75:27–35.

Foster NEV, Zatorre RJ. 2009. A role for the intraparietal sulcus in transforming musical pitch information. *Cereb Cortex*. 20:1350–1359.

Golestani AM, Faraji-Dana Z, Kayvanrad M, Setsompop K, Graham SJ, Chen JJ. 2018. Simultaneous multislice resting-state functional magnetic resonance imaging at 3 tesla: slice-acceleration-related biases in physiological effects. *Brain Connect*. 8:82–93.

Griffiths TD. 2000. Musical hallucinosis in acquired deafness. Phenomenology and brain substrate. *Brain*. 123:2065–2076.

Gu J, Zhang H, Liu B, Li X, Wang P, Wang B. 2019. An investigation of the neural association between auditory imagery and perception of complex sounds. *Brain Struct Funct*. 224:2925–2937.

Halpern A. 2001. Cerebral substrates of musical imagery. *Ann NY Acad Sci*. 930:179–192.

Halpern AR. 2015. Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*. 25:37–47.

Halpern AR, Zatorre RJ. 1999. When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb Cortex*. 9:697–704.

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. 2004. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*. 42:1281–1292.

Hamilton LS, Oganian Y, Chang EF. 2020. Topography of speech-related acoustic and phonological feature encoding throughout the human core and parabelt auditory cortex. *bioRxiv*. doi: 10.1101/2020.06.08.121624.

Hardyck CD, Petrinovich LF. 1970. Subvocal speech and comprehension level as a function of the difficulty level of Reading material. *J Verbal Learn Verbal Behav*. 9:647–652.

Hasson U, Malach R, Heeger DJ. 2009. Reliability of cortical activity during natural stimulation. *Trends Cogn Sci*. 14:40–48.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science*. 303:1634–1640.

Herholz SC, Halpern AR, Zatorre RJ. 2012. Neuronal correlates of perception, imagery, and memory for familiar tunes. *J Cogn Neurosci*. 24:1382–1397.

Herholz SC, Lappe C, Knief A, Pantev C. 2008. Neural basis of music imagery and the effect of musical expertise. *Eur J Neurosci*. 28:2352–2360.

Hickok G, Buchsbaum BR, Humphries C, Muftuler LT. 2003. Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci*. 15:673–682.

Hickok G, Houde J, Rong F. 2011. Sensorimotor integration in speech processing computational basis ans neural organization. *Neuron*. 69:407–422.

Honey CJ, Thompson CR, Lerner Y, Hasson U. 2012. Not lost in translation: neural responses shared across languages. *J Neurosci*. 32:15277–15283.

Hubbard TL. 2010. Auditory imagery: empirical findings. *Psychophysiology*. 136:302–329.

Hubbard TL. 2018. Some methodological and conceptual considerations in studies of auditory imagery. *Auditory Perception & Cognition*. 1:6–41.

Ikeda S, Shibata T, Nakano N, Okada R, Tsuyuguchi N, Ikeda K, Kato A. 2014. Neural decoding of single vowels during covert articulation using electrocorticography. *Front Hum Neurosci*. 8:1–8.

Jacobsen J-H, Stelzer J, Fritz TH, Chételat G, La Joie R, Turner R. 2015. Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*. 138:2438–2450.

James W. 1890. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.

Keogh R, Pearson J. 2011. Mental imagery and visual working memory. *PLoS One*. 6:e29221.

Kleiner M, Brainard DH, Pelli DG, Ingling A, Murray RM. 2007. What's new in Psychtoolbox-3. *Perception*. 36:1–16.

Kotz SA, Schwartze M, Schmidt-Kassow M. 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex*. 45:982–990.

Kraemer DJM, Macrae CN, Green AE, Kelley WM. 2005. Sound of silence activates auditory cortex. *Nature*. 434:158.

Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:1–28.

Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP. 2009. Brain activation during anticipation of sound sequences. *J Neurosci*. 29:2477–2485.

Lerner Y, Honey CJ, Silbert LJ, Hasson U. 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci*. 31:2906–2915.

Li Y, Luo H, Tian X. 2020. Mental operations in rhythm: motor-to- sensory transformation mediates imagined singing. *PLoS Biol*. 18:e3000504.

Lima CF, Krishnan S, Scott SK. 2016. Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends Neurosci*. 39:527–542.

Linke AC, Cusack R. 2015. Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. *J Cogn Neurosci*. 27:1322–1333.

Lucas BJ, Schubert E, Halpern AR. 2010. Perception of emotion in sounded and imagined music. *Music Percept*. 27: 399–412.

Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone NE, Riegar J, Schalk G, Knight RT, Pasley BN. 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front Neuroeng*. 7:1–15.

Martin S, Mikutta C, Leonard MK, Hungate D, Koelsch S, Shamma S, Chang EF, Millán JR, Knight RT, Pasley BN. 2018. Neural encoding of auditory features during music perception and imagery. *Cereb Cortex*. 28:4222–4233.

McCullough Campbell S, Margulis EH. 2015. Catching an earworm through movement. *J New Music Res*. 44:347–358.

McGuigan FJ, Keller B, Stanton E. 1964. Covert language responses during silent Reading. *J Educ Psychol*. 55: 339–343.

Meyer K, Kaplan JT, Essex R, Webber C, Damasio H, Damasio A. 2010. Predicting visual stimuli on the basis of activity in auditory cortices. *Nat Neurosci*. 13:667–668.

Meyer M, Elmer S, Baumann S, Jancke L. 2007. Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restor Neurol Neurosci*. 25:411–431.

Mikumo M. 1994. Motor encoding strategy for pitches of melody. *Music Percept*. 12:175–197.

Morillon B, Baillet S. 2017. Motor origin of temporal predictions in auditory attention. *PNAS*. 114:E8913–E8921.

Morillon B, Hackett TA, Kajikawa Y, Schroeder CE. 2015. Predictive motor control of sensory dynamics in auditory active sensing. *Curr Opin Neurobiol*. 31:230–238.

Morillon B, Schroeder CE, Wyart V. 2014. Motor contributions to the temporal precision of auditory attention. *Nat Commun*. 5:1–9.

Moulton ST, Kosslyn SM. 2009. Imagining predictions: mental imagery as mental emulation. *Philos Trans R Soc Biol Sci*. 364:1273–1280.

Musch K, Himberger K, Valiante TA, Honey CJ. 2019. Transformation of speech sequences in human sensorimotor circuits. *PNAS*. 117:3203–3213.

Nguyen M, Vanderwal T, Hasson U. 2019. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*. 184:161–170.

Nichols TE, Holmes AP. 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 15:1–25.

Norman-Haignere S, Kanwisher NG, McDermott JH. 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*. 88: 1281–1296.

Nozaradan S, Schönwiesner M, Caron-Desrochers L, Lehmann A. 2016. Enhanced brainstem and cortical encoding of sound during synchronized movement. *NeuroImage*. 142: 231–240.

Oh J, Kwon JH, Yang PS, Jeong J. 2013. Auditory imagery modulates frequency-specific areas in the human auditory cortex. *J Cogn Neurosci*. 25:175–187.

Okada K, Matchin W, Hickok G. 2018. Neural evidence for predictive coding in auditory cortex during speech production. *Psychon Bull Rev*. 25:423–430.

Palmiero M, Cardi V, Belardinelli MO. 2011. The role of vividness of visual mental imagery on different dimensions of creativity. *Creat Res J*. 23:372–375.

Pei X, Barbour DL, Leuthardt EC, Schalk G. 2011. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng*. 8: 1–11.

Pelli DG. 1997. The VideoToolbox software for visual psychophysics transforming numbers into movies. *Spat Vis*. 10:437–442.

Peretz I, Gosselin N, Belin P, Zatorre RJ, Plailly J, Tillmann B. 2009. Music lexical networks: the cortical Organization of Music Recognition. *Ann N Y Acad Sci*. 1169:256–265.

Rampinini AC, Handjaras G, Leo A, Cecchetti L, Ricciardi E, Marotta G, Pietrini P. 2017. Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. *Sci Rep*. 7:1–13.

Rauschecker JP. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res*. 271:16–25.

Rauschecker JP. 2018. Where, when, and how: are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*. 98:262–268.

Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*. 12:718–724.

Regev M, Honey CJ, Simony E, Hasson U. 2013. Selective and invariant neural responses to spoken and written narratives. *J Neurosci*. 33:15978–15988.

Regev M, Simony E, Lee K, Tan KM, Chen J, Hasson U. 2019. Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cereb Cortex*. 29:4017–4034.

Repp BH, Su Y-H. 2013. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon Bull Rev*. 20: 403–452.

Reznik D, Mukamel R. 2019. Motor output, neural states and auditory perception. *Neurosci Biobehav Rev*. 96:116–126.

Saur D, Kreher BW, Schnell S, Kummerer D, Kellmeyer P, Vry M-S, Umarova R, Musso M, Glauche V, Abel S, et al. 2008. Ventral and dorsal pathways for language. *PNAS*. 105: 18035–18040.

Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK. 2012. The future of memory: remembering, imagining, and the brain. *Neuron*. 76:677–694.

Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BTT. 2018. Local-global Parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb Cortex*. 28:3095–3114.

Scheel N, Chang C, Mamlouk AM. 2014. The importance of physiological noise regression in high temporal resolution fMRI. *Artificial Neural Networks and Machine Learning – ICANN*. 2014(8681):829–836.

Schön D, Tillmann B. 2015. Short- and long-term rhythmic interventions: perspectives for language rehabilitation. *Ann N Y Acad Sci*. 1337:32–39.

Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. 2010. Dynamics of active sensing and perceptual selection. *Curr Opin Neurobiol*. 20:172–176.

Schubotz RI. 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn Sci*. 11:211–218.

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *PNAS*. 111: E4687–E4696.

Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U. 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun*. 7:1–13.

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline J-B, LeBihan D, Dehaene S. 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage*. 33:1104–1116.

Tian X, Zarate JM, Poeppel D. 2016. Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*. 77:1–12.

Vetter P, Smith FW, Muckli L. 2014. Decoding sound and imagery content in early visual cortex. *Curr Biol*. 24:1256–1262.

Weir G, Williamson VJ, Müllensiefen D. 2015. Increased involuntary musical mental activity is not associated with more accurate voluntary musical imagery. *Psychomusicology: Music, Mind, and Brain*. 25:48–57.

Wolpert DM, Ghahramani Z, Jorden MI. 1995. An internal model for sensorimotor integration. *Science*. 269:1880–1882.

Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X. 2009. Functional maps of human auditory cortex: effects of acoustic features and attention. *PLoS One*. 4:e5183.

Yao B, Belin P, Scheepers C. 2011. Silent Reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *J Cogn Neurosci*. 23:3146–3152.

Yoo S-S, Lee CU, Choi BG. 2001. Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport*. 12:3045–3049.

Zadbood A, Chen J, Leong YC, Norman KA, Hasson U. 2017. How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb Cortex*. 27:4988–5000.

Zarahn E, Aguirre GK, D'Esposito M. 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage*. 5: 179–197.

Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*. 6: 37–46.

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews*. 8:547–558.

Zatorre RJ, Halpern A. 1993. Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*. 31:221–232.

Zatorre RJ, Halpern AR, Perry DW, Meyer E, Evans AC. 1996. Hearing in the Mind's ear: a PET investigation of musical imagery and perception. *J Cogn Neurosci*. 8: 29–46.

Zhang W, Liu Y, Wang X, Tian X. 2020. The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cogn Neurosci*. 11:194–204.